

**Universidad Autónoma de Madrid**  
**Departamento de Lingüística General, Lenguas Modernas,**  
**Lógica y Filosofía de la Ciencia, Teoría de la Literatura y**  
**Literatura Comparada**  
**Laboratorio de Lingüística Informática**

**Compilación de un corpus de habla espontánea de chino**  
***putonghua* para su aplicación en la enseñanza como**  
**lengua segunda a hispanohablantes**

**DONG Yang**

**Tesis doctoral dirigida por el Dr. Antonio Moreno Sandoval**

**2011**

## **Agradecimientos**

En primer lugar, quisiera dar las gracias al director de esta tesis, Dr. Antonio Moreno Sandoval, especialmente por aceptarme para realizar esta tesis bajo su dirección. Además, quisiera agradecer la confianza y paciencia que él ha depositado en este trabajo, así como su constante estímulo, sus comentarios y consejos tan valiosos para mi tesis.

Gracias a la beca concedida por la Agencia Española de Cooperación Internacional, he podido seguir mis estudios del Programa de Doctorado en la Facultad de Filosofía y Letras de la Universidad Autónoma de Madrid y dedicarme totalmente a la tesis.

A la Dra. Taciana Fisac, coordinadora del Programa de Doctorado “España y Latinoamérica Contemporáneas” y catedrática del Centro de Estudios de Asia Oriental de la Universidad Autónoma de Madrid, quiero extenderle un sincero agradecimiento por su disponibilidad, generosidad y ayuda incondicional en todo momento. Fue ella quien organizó y promovió este Programa de Doctorado de la Universidad Autónoma de Madrid, con la colaboración de la Universidad de Lenguas Extranjeras de Beijing.

Quisiera agradecer a los miembros del Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid por la ayuda que me han ofrecido durante estos años de aprendizaje. Se trata de un soporte profesional muy importante. Muchas gracias por permitirme vivir una experiencia tan inovidable y bonita para mi formación.

Muchas gracias al Dr. José María Guirao de la Universidad de Granada que, con su gran profesionalidad en el campo de la informática, siempre me ha prestado ayuda para resolver todos los problemas técnicos.

Un agradecimiento especial al Sr. Wang Zhiwei, director de la Oficina de Asuntos Educativos de la Embajada de la República

Popular China en España, por los ánimos y los datos relacionados con la enseñanza de chino en España que me ha facilitado.

Tengo la suerte de haber conocido a la familia de Marisol Cerezo y de José Ramón. A esta familia le agradezco especialmente su amabilidad, amistad y ánimo durante mi estancia en España. Los considero como mi familia en España. Con ellos he compartido muchos momentos que siempre llevaré en mi corazón y gracias a los cuales nunca me he sentido sola en España.

Estoy profundamente agradecida a todos los participantes del corpus y la Universidad de Estudios Internacionales de Beijing que es mi lugar de trabajo, por su comprensión, colaboración y apoyo a la enseñanza de idiomas. Han contribuido mucho a la elaboración de este corpus de habla espontánea de chino *putonghua*. Sin ellos, esta tesis no habría podido ver la luz.

Finalmente, no me olvido de mis amigos, alumnos y familiares. Muchas gracias por todo. Os quiero a todos. Una mención especial va para mi marido An. Soy afortunada por contar siempre con su amor, comprensión y ánimo que me sirven como fortaleza firme para seguir adelante.

## Índice general

<b>1. Introducción</b>	<b>8</b>
1.1 Motivos del trabajo	8
1.2 Objetivos de la investigación	10
1.3 Estructura de la tesis	11
<b>PARTE PRIMERA</b>	<b>13</b>
<b>Bases teóricas</b>	<b>13</b>
<b>2. La Lingüística de Corpus</b>	<b>14</b>
2.1 Aspectos generales	14
2.2 El corpus C-ORAL-ROM	29
2.3. La Lingüística de Corpus en China	37
<b>3. El uso de corpus en la enseñanza de lenguas</b>	<b>41</b>
3.1 Las aplicaciones de corpus en la enseñanza de lenguas	41
3.2 La experiencia del laboratorio LLI-UAM	47
<b>4. El chino putonghua</b>	<b>57</b>
4.1 Visión global	57
4.2 Evolución del idioma chino	58
4.3 Caracteres chinos	62
4.4 Hanyu pinyin	66
4.5 Dialectos	69
4.6 Lengua oral y escrita	70
4.7 Gramática	74
4.8 Léxico	81
4.9 La enseñanza de la lengua china como lengua extranjera en España	83
<b>PARTE SEGUNDA</b>	<b>87</b>
<b>La aplicación</b>	<b>87</b>
<b>5. Diseño del corpus</b>	<b>88</b>
5.1 Aspectos generales	88
5.2 Elección de los participantes	91

5.3 Tres tipos de grabación-----	94
5.4 Comparación con el corpus CADCC-----	96
5.5 Diseño final-----	107
<b>6. Metodología-----</b>	<b>108</b>
6.1 Grabación y digitalización-----	108
6.2 Transcripción ortográfica-----	111
6.3 Alineamiento-----	124
6.4 Transcripción de pinyin-----	127
6.5 Conversión XML-----	127
6.6 Resumen-----	130
<b>7. Aplicación a la enseñanza de lengua segunda mediante el desarrollo de recursos basados en el corpus: propuestas didácticas-----</b>	<b>133</b>
7.1 El uso de corpus en la enseñanza del léxico-----	134
7.2 El uso de corpus en la enseñanza de la gramática-----	144
7.3 El uso de corpus en la enseñanza de comprensión auditiva-----	172
7.4 El uso de corpus en la enseñanza de fonemas-----	175
<b>8. Conclusión-----</b>	<b>182</b>
<b>Bibliografía-----</b>	<b>185</b>
<b>Apéndice A-----</b>	<b>193</b>
<b>Apéndice B-----</b>	<b>241</b>
<b>Apéndice C-----</b>	<b>249</b>

## **Índice de tablas y figuras**

### ***Parte primera***

Tabla 2.1 Matriz del diseño de C-ORAL-ROM_____	32
Tabla 2.2 Tabla general del corpus multimedia C-ORAL-ROM	35
Figura 3.1 Aplicaciones de Corpus en la Enseñanza de Lenguas_____	42
Tabla 3.2 Contenidos gramaticales_____	49
Tabla 3.3 Nociones y funciones comunicativas_____	50
Tabla 3.4 Contenidos léxicos_____	51
Figura 3.5 Herramienta de búsqueda de C-ORAL-ROM_____	52
Tabla 4.1 Número de caracteres en diccionarios chinos_____	62
Tabla 4.2 Contabilidad de número de sílabas de la lengua oral y la escrita_____	72
Tabla 4.3 Contabilidad de número de morfemas de la lengua oral y la escrita____	72
Tabla 4.4 Contabilidad de número de significados de la lengua oral y la escrita__	73
Tabla 4.5 Contabilidad de palabras que llevan tres sufijos habituales en dialectos, lengua oral y lengua escrita _____	73
Tabla 4.6 Ejemplos de relaciones entre categorías y elementos sintácticos_____	78

### ***Parte Segunda***

Tabla 5.1 Edad de los participantes de C-ORAL-CHINA_____	92
Tabla 5.2 Parámetros en el diseño del corpus C-ORAL-CHINA_____	95
Tabla 5.3 Datos básicos del corpus CADCC_____	97
Tabla 5.4 Distribución de tiempo/tema de cada pareja de interlocutores_____	98
Tabla 5.5 Velocidad caracteres / segundo del subcorpus Formal en contexto público._____	100
Tabla 5.6 Velocidad caracteres / segundo del subcorpus Informal_____	101
Tabla 5.7 Velocidad caracteres / segundo del subcorpus de	102

Medios de comunicación_____	
Figura 5.8 Ejemplo de transcripción en CADCC_____	103
Figura 5.9 Ejemplo de transcripción en C-ORAL-CHINA_____	104
Figura 5.10 Ejemplo de transcripción de <i>pinyin</i> (sin segmentación de palabras)_____	105
Figura 5.11 Ejemplo de transcripción de <i>pinyin</i> (con segmentación de palabras)_____	105
Tabla 5.12 Convenciones de transcripción de signos paralingüísticos de CADCC_____	106
Figura 6.1 Modelo de permiso de grabación_____	109
Figura 6.2 Edición del sonido con Cool Edit_____	110
Figura 6.3 Programa de transcripción Transana_____	125
Figura 6.4 Alineamiento en Transana_____	126
Figura 6.5 Salida del alineamiento en transana_____	128
Figura 6.6 Ejemplo de salida a formato XML_____	130
Tabla 6.7 Convenciones de la transcripción_____	132
Figura 7.1 Ejemplo de C-ORAL-CHINA_____	137
Figura 7.2 Pantalla principal de la página web_____	138
Figura 7.3 Buscador de C-ORAL-CHINA_____	139
Figura 7.4 Ejemplos de la palabra buscada “感动”_____	139
Figura 7.5 Ejemplos de la palabra buscada “激动”_____	140
Figura 7.6 Ficheros de <i>Transcripciones</i> _____	141
Figura 7.7 Contenido del fichero chin02_____	142
Figura 7.8 Informaciones resumidas de todos los ficheros de C-ORAL-CHINA	142
Tabla 7.9 Ejemplos de la palabra buscada “采用”_____	143
Figura 7.10 Ejemplos de la palabra buscada “采纳”_____	144
Figura 7.11 Ejemplo de supermercado de C-ORAL-CHINA_____	174
Tabla 7.12 Datos básicos del subcorpus de <i>pinyin</i> _____	176
Tabla 7.13 Frecuencia de consonante inicial_____	177
Tabla 7.14 Frecuencia de vocal según el tono_____	179
Tabla 7.15 Frecuencia y porcentaje total de vocal_____	181

## 1. Introducción

### *1.1 Motivos del trabajo*

Hoy en día nos encontramos inmersos en la llamada “era de la informática”. Diferentes ámbitos de la investigación lingüística están influidos por el gran impacto tecnológico y el peculiar tratamiento informático de nuestros días. La lingüística de corpus se cuenta entre ellos.

El primer corpus moderno de inglés, legible por máquina, el *Brown corpus* (una compilación de inglés estadounidense de aproximadamente un millón de palabras), fue creado en los años sesenta del siglo pasado (McEnery et al., 2006). A partir de los años 80, el número y tamaño de los corpus y los estudios basados en ellos han crecido espectacularmente. La lingüística de corpus ha entrado desde entonces en un crecimiento progresivo. La explotación constante de los programas y *softwares* informáticos no sólo ha impulsado la construcción de corpus, sino que ha aumentado además la capacidad de tratamiento de éstos. La gran cantidad de corpus anotados han favorecido la investigación lingüística. Gracias a todo esto, se ha ido ganando paulatinamente en métodos de investigación más avanzados.

En la actualidad, los corpus en soporte informático constituyen una herramienta de gran utilidad para el desarrollo de la lingüística. Según la necesidad concreta de los investigadores, la creación de corpus puede tener diversos fines, tales como la elaboración de diccionarios, el estudio de frecuencia de palabras en un determinado dominio comunicativo, la investigación de algunos fenómenos lingüísticos, la enseñanza de lenguas, etc. En suma, la lingüística de corpus se puede considerar como una metodología con gran rango de aplicación en muchas áreas y teorías de la lingüística (McEnery et al., 2006).



El Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (LLI-UAM) cuenta con muchos años de experiencia en la elaboración y aplicación de corpus. Algunos de sus recursos lingüísticos son: Corpus Oral de Referencia de la Lengua Española Contemporánea, Corpus de Referencia de la Lengua Española en la Argentina, Corpus de Referencia de la Lengua Española en Chile, *Spanish Treebank Corpus*, C-ORAL-ROM, Corpus de Habla Infantil Espontánea del Español, Corpus Árabe-Español, etc. Entre ellos, C-ORAL-ROM es un corpus multilingüe que se caracteriza por su espontaneidad: los textos se han grabado en su contexto real y sin guión preestablecido.

En esta tesis, aprovechando la experiencia y la metodología del LLI-UAM, se estudia la compilación de un corpus de habla espontánea de chino *putonghua* para su aplicación en la enseñanza como segunda lengua a hispanohablantes.

La lengua no sólo es una herramienta de comunicación, sino que también sirve como un puente que impulsa el conocimiento entre gentes de diferentes procedencias y referencias culturales. A medida que se consolidan las relaciones económicas, políticas, culturales y educativas entre China, España y América Latina, el interés por el aprendizaje del idioma chino está creciendo de manera acelerada en el mundo hispanohablante. No obstante, en comparación con otros países, todavía queda mucho por explorar en lo que a la enseñanza de la lengua china a los hispanohablantes se refiere.

La combinación de corpus y enseñanza de lenguas se ha convertido en una metodología merecedora de una atención especial por lo mucho que puede aportar. Por citar sólo algunas de las muchas aplicaciones posibles en la polifacética enseñanza de lenguas, los corpus pueden usarse, por ejemplo, en la preparación de manuales, en el uso de DDL (*Data driven Learning*), en la evaluación del nivel del dominio de la lengua, la formación de profesores, el aprendizaje de una lengua para un uso específico, etc. En nuestro caso, creemos que un corpus de habla espontánea de chino *putonghua* elaborado desde la perspectiva

del aprendiz hispanohablante de chino podría ser un instrumento eficaz tanto para profesores como para alumnos. Por medio de las herramientas concretas en cada caso, todos los interesados pueden aprovechar los recursos del corpus para el estudio, según el nivel, la necesidad y el interés de cada uno.

Dolores Font atribuye una importancia crucial a la investigación a partir del habla espontánea, “[...] porque es el modelo que nos proporciona el habla real y genuina, la lengua en su estado natural (Font, 2007: 318)”.

Massimo Moneglia resalta su necesidad:

It is important to achieve knowledge on spoken language for the purpose of the development of linguistic engineering, which, in the era of digital communication, must address the primary role played by spontaneous speech within natural communication (Cresti y Moneglia, 2005: 1).

Un corpus de habla espontánea de chino puede constituir una fuente ideal de datos vivos en situaciones comunicativas de uso real, lo cual es muy útil para que los estudiantes puedan afrontar situaciones comunicativas cotidianas básicas en una eventual estancia en China. Puede servir como manual moderno y suplementario para el aprendizaje de idiomas, compensando las deficiencias de los materiales tradicionales. Puede además aportar conocimientos de la cultura de la lengua y proporcionar datos estadísticos necesarios, etc.

## ***1.2 Objetivos de la investigación***

Los objetivos de este estudio se pueden resumir en 5 puntos diferenciados:

1. Recogida de muestras de chino *putonghua* espontáneo para la creación de un corpus con alrededor de 10 horas de grabación.

2. Transcripción ortográfica de las muestras.
3. Alineamiento de la transcripción y el sonido de todo el corpus.
4. Confección de una página web para la consulta del corpus.
5. Aplicación a la enseñanza de lengua segunda mediante el desarrollo de recursos basados en el corpus.

### ***1.3 Estructura de la tesis***

Con la finalidad de dar cumplimiento a los objetivos arriba mencionados, la tesis está organizada en dos partes: la parte teórica y la práctica. En la primera parte, se enuncia la teoría sobre la que se fundamenta la creación del corpus. En la segunda, se expone la compilación del corpus y se estudia su aplicación a la enseñanza.

La primera parte consta de tres capítulos. En el segundo capítulo, bajo el primer apartado, se presenta una amplia revisión de la lingüística de corpus y se resumen las características generales de un corpus. El segundo se dedica a la presentación del corpus C-ORAL-ROM como modelo para el desarrollo de nuestro proyecto. En el último, se enumeran los corpus actuales en China y se resaltan las características entre uno y otro.

El tercer capítulo trata del uso de corpus en la enseñanza de lenguas. Con el desarrollo de la lingüística de corpus, cada vez hay más estudios sobre su aplicación a la enseñanza. Aquí se resumen los antecedentes históricos a nivel mundial desde un punto de vista histórico y se presenta la experiencia del LLI-UAM que durante varios años ha estado dedicado a la explotación de los corpus elaborados.

El cuarto capítulo se dedica al chino *putonghua*. Aparte de una visión global, se explican su origen, evolución, escritura, *hanyu pinyin*, léxico, lengua oral y escrita, gramática, etc. A continuación, se describe el desarrollo de la enseñanza de la lengua china como lengua extranjera en España, tomándolo como referencia de la situación de la enseñanza de chino a hispanohablantes.

Con el capítulo quinto comienza la parte de la tesis elaborada con vistas a su aplicación como material de enseñanza lingüística. Se trata, ante todo, del diseño del corpus. Al hacer la elección de los participantes, hay que tener en cuenta que China es un país con muchos dialectos y el objetivo de este corpus es el de servir de material de enseñanza. El corpus cuenta con grabaciones de monólogos y diálogos que responden a tres diferentes tipos de registro verbal: formal en contexto público, medios de comunicación e informal. Como hoy por hoy ya existen varios corpus de chino *putonghua* hechos por diferentes organismos, finalmente, se hace una comparación del nuestro con uno de referencia: el corpus CADCC.

En el capítulo sexto se indica la metodología que abarca grabación, digitalización, transcripción, revisión, alineamiento, conversión a lenguaje computacional XML y transcripción de *pinyin*. Como hemos comentado antes, la construcción del corpus de habla espontánea de chino *putonghua* se basa en la metodología empleada para la elaboración de C-ORAL-ROM y aprovecha la experiencia y los recursos del LLI-UAM.

A continuación, en el siguiente capítulo, se estudia cómo se aprovechan los recursos basados en un corpus para aplicarlos a la enseñanza de lenguas planteando propuestas didácticas propias. Se divide en cuatro subcapítulos: el uso de corpus en la enseñanza del léxico, de la gramática, de comprensión auditiva y el estudio de los fonemas en el corpus, basado en la transcripción de *pinyin*.

Finalmente, se ofrece un capítulo conclusivo en donde, partiendo de las observaciones y las experiencias obtenidas en la realización del trabajo, se proponen una serie de tareas para la investigación en el futuro.

Después de los capítulos, se registra la bibliografía en la que se listan los libros consultados y las páginas web visitadas. Para concluir este trabajo, se han incluido tres apéndices donde se ofrecen ejemplos de transcripción y diferentes tablas y figuras con datos estadísticos.

---

## **PARTE PRIMERA**

### **Bases teóricas**

## **2. La Lingüística de Corpus**

La Lingüística de Corpus es una subdisciplina de la lingüística, descrita como “the study of language on the basis of text corpora (Aijmer y Altenberg, 1991: 29)”. A lo largo de los últimos años, gracias a la creciente capacidad tecnológico-computacional, la Lingüística de Corpus ha logrado un gran progreso.

Uno de los temas en disputa en la comunidad científica es si la Lingüística de Corpus es una metodología empírica de trabajo o más bien una rama independiente de la lingüística. Diferentes estudiosos de la cuestión han expresado sus propias opiniones. En esta tesis, no trataremos de profundizar en ese debate sino que vamos a analizar y comentar su aplicación porque, sea como sea, ninguna escuela científica puede negar su importancia y papel en la investigación lingüística.

Para nuestros antepasados hubiese sido imposible imaginar cómo hoy día seríamos capaces de registrar y manipular recursos de centenares de millones de palabras. Según la necesidad, por medio de herramientas como analizadores sintácticos, etiquetadores morfológicos o programas de concordancias, los investigadores pueden desde hacer la anotación de un corpus determinado, hasta sacar datos estadísticos como la frecuencia o la probabilidad para hacer un análisis cuantitativo, pasando por describir las características de la lengua, aplicar esos mismos datos en la enseñanza de lenguas o verificar alguna hipótesis respecto a éstas, etc. En suma, su utilidad es muy amplia de áreas de la investigación lingüística.

### ***2.1 Aspectos generales***

### 2.1.1 Definición de corpus

La palabra “corpus” es de origen latino y significa “cuerpo”. Según el Diccionario de la Real Academia Española de la Lengua, se trata de todo “conjunto lo más extenso y ordenado posible de datos o textos científicos, literarios, etc., que pueden servir de base a una investigación”.

En los años sesenta del siglo pasado, el primer corpus legible por máquina, *The Brown corpus*, fue compilado en los Estados Unidos, lo que significó el nacimiento de los corpus informatizados modernos. A partir de ese momento, el concepto “corpus” no ha parado de evolucionar, reclamando nuevos significados para sí en la lingüística moderna. Diversas definiciones tratan de dar cuenta de ello:

A collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information.

A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language (McEnery et al., 2006: 4).

Un corpus es un conjunto de textos de lenguaje natural e irrestricto, almacenados en un formato electrónico homogéneo, y seleccionados y ordenados, de acuerdo con criterios explícitos, para ser utilizados como modelo de un estado o nivel de lengua determinado, en estudios o aplicaciones relacionados en mayor o menor medida con el análisis lingüístico (Santalla del Río, 2005: 45-46).

En sentido general, se puede entender que los corpus modernos se caracterizan por los siguientes rasgos fundamentales:

**Autenticidad de los datos:** Las muestras recopiladas se recogen desde textos auténticos en situaciones reales, reflejando el uso de la lengua.

**Legible por máquina:** Los textos se hallan disponibles en un formato electrónico. Se codifican, guardan y ordenan mediante la tecnología informática, facilitando así el análisis y el procesamiento de los datos.

**Anotación:** Algunos estudiosos abogan por la anotación, opinando que cuenta con ventajas como la facilidad en la explotación y uso de los materiales, la diversidad en la función, la claridad en el análisis, etc. En contraste, los críticos creen que con la involucración subjetiva del investigador que anota, el material pierde objetividad y puede ser difícil garantizar la fiabilidad de la anotación. En la práctica, la anotación tiende a ser cada vez más exhaustiva. Para cada corpus concreto, dependiendo del objetivo, se puede buscar una manera apropiada de anotación.

### ***2.1.2 Criterios en el diseño de corpus.***

A continuación, se mencionan ciertos criterios que se deben considerar durante el diseño de cualquier corpus.

**Objetivo:** La creación de corpus puede servir tanto para uso personal como para la investigación lingüística. Hay desde luego muchos objetivos: redacción de diccionarios, contabilidad de frecuencia, análisis de algún fenómeno lingüístico, estudio de la lengua de algún escritor, enseñanza de lenguas, traducción automática, etc. El objetivo decide el ámbito y tamaño de la selección de muestras.

**Representatividad:** Con este concepto se resalta que el corpus debe poseer la cualidad de mostrar fidedignamente, a través de sus contenidos, la realidad de la lengua en la gran comunidad de hablantes que la utilizan. Dicha representatividad se destaca con el fin de lograr la fiabilidad de la deducción sacada sobre la base del análisis del corpus. Sin embargo, es imposible llegar a una representatividad absoluta. Por lo general, se recomienda que los



diseñadores traten de seleccionar muestras que puedan garantizar la variedad de la lengua de estudio para que las muestras limitadas puedan reflejar las características de los fenómenos lingüísticos de la lengua infinita lo mayor posible. Sobre todo, si se quiere construir un corpus general para hacer la descripción de alguna lengua en su conjunto, la representatividad constituye una condición imprescindible. Un corpus amplio, completo y equilibrado es ciertamente muy útil para la investigación y análisis. Los corpus famosos como *The Brown Corpus* y BNC son de este tipo.

**Equilibrio:** Si un corpus se destina a estudiar las variedades de una lengua, hay que determinar con antelación las que estarán representadas y calcular su proporción en el corpus, de manera que nos permitan validar aquellas conclusiones extraídas. Por ejemplo, si se quiere elaborar un corpus de dialectos del chino, primero hay que analizar los dialectos principales de China, y luego decidir en qué porcentaje habrá de recoger las muestras de distintas zonas de China. Otro ejemplo es el corpus diacrónico CORDE (Corpus Diacrónico del Español). Abarca tres épocas fundamentales: la Edad Media, el Siglo de Oro y la época Contemporánea, además de una geografía hispanohablante muy amplia. Aquí el diseño previo de la estructura interna entre diferentes textos es crucial.

**Tamaño:** Siempre ha sido cuestión discutida el tamaño de un corpus. No hay un estándar de cuántas palabras son adecuadas para un corpus concreto. Hay dos tendencias en la compilación de corpus (Flowerdew, 2002). De un lado, los corpus se tornan cada vez más grandes, dadas la manipulación y explotación informática de gran capacidad. De otro lado, se han compilado muchos corpus especializados que son más pequeños. Susan Hunston lo comenta en su libro *Corpora in Applied Linguistics*:

[...] the Bank of English is currently about 400 million. [...] In the Bank of English corpus, for example, a search for the word point gives almost 143,000 hits. Few researchers can obtain useful information simply by looking at so many concordance lines. One solution is to use a smaller corpus, which will give less data. Another is to keep the large corpus,

but to use software which will make a selection of data from the whole, [...] (Hunston, 2006: 25).

En nuestra opinión, como hay tantos tipos de corpus y diferentes funciones, la extensión de un corpus debe depender de la finalidad para la que se cree. Obviamente, un corpus que recopila extensos textos con el fin de servir como muestra representativa de una lengua para reflejar su uso se distingue mucho de un corpus confeccionado para algún estudio específico. Lo más importante es que la extensión del corpus pueda satisfacer la necesidad del investigador.

Hasta la fecha, la discusión de estos y otros aspectos sigue abierta. En cualquier caso, se trata siempre de criterios y factores estrechamente relacionados con la finalidad de un corpus determinado. En el capítulo 5 abordaremos de nuevo estos criterios que dirigen la elaboración de un corpus, junto a la práctica de nuestra compilación de C-ORAL-CHINA.

### ***2.1.3 Evolución histórica de la idea de corpus***

La historia de los corpus que concierne al estudio de lenguas y a la búsqueda de palabras y frases en múltiples contextos puede remontarse al siglo XIII. Los eruditos medievales y su cohorte escolástica estudiaron la Biblia página por página y registraron manualmente en índices las palabras de cada línea (McCarthy y O’Keeffe, 2010: 3).

Gracias al progreso de la tecnología informática, la eficacia y el tamaño de la compilación de corpus han elevado mucho sus cotas, contribuyendo firmemente a un uso más amplio. Podemos dividir la historia de su desarrollo en dos épocas.

La época con anterioridad a la aparición del ordenador es la de los corpus tradicionales (*corpus on slips of paper*). Los materiales se coleccionaban y organizaban manualmente. Como no tenían formato electrónico, los datos se registraban en papel.

Por ejemplo, el diccionario de inglés del Dr. Samuel Johnson, publicado en 1755, fue el resultado de muchos años de trabajo. Se apuntaba en gran cantidad de papel el uso de las palabras inglesas durante el período desde 1560 hasta 1660. Otro ejemplo es *The Oxford English Dictionary* (OED) en la década de los 80 del siglo XIX. Sin esos más de tres millones de fichas que le sirvieron de base, este famoso diccionario no habría sido compilado. La experiencia demuestra que la recogida de datos es esencial para el trabajo de los estudiosos.

Como Leech señala, la década de los 50 es la era de los estructuralistas estadounidenses, entre los cuales figuran Harris, Fries y Hill, etc (McCarthy y O’Keeffe, 2010). Fueron ellos los precursores de los corpus, no sólo en el sentido de la recolección de datos de lengua usados en la vida real, sino también en la práctica de asignarles un puesto central del estudio lingüístico.

Sin embargo, a finales de los años 50, la publicación de los libros del padre de la gramática transformacional y generativista, Noam Chomsky provocó un cambio radical de enfoque en los estudios lingüísticos, haciendo que la investigación de corpus fuera momentáneamente apartada y se viese muy afectada.

Chomsky atribuye gran importancia a la intuición del hablante, según la opinión de que la introspección es un criterio suficiente para determinar la competencia y el empleo adecuado de una lengua. En otras palabras, la introspección permitiría determinar la gramaticalidad de un enunciado o resolver ambigüedades. Según él, el principal objetivo de la investigación lingüística es establecer un modelo de reconocimiento de la lengua que pueda reflejar la realidad psicológica y la competencia lingüística del hablante. Esencialmente, los corpus son recopilaciones de palabras, sujetas a errores y variaciones. Resulta inevitable que cualquier corpus sea incompleto y finito. No se puede coleccionar todo. La investigación basada en este modelo de empirismo se limitaría, pues, a explicar una parte de la competencia y no sería capaz de reflejar la infinita capacidad generativa de las lenguas naturales. Por todo esto, este lingüista postula que los corpus no pueden considerarse como una herramienta relevante para los lingüistas (Ding Xinshan, 2005).

Durante los años 60 y 70, predominó el racionalismo basado en la intuición del lingüista. Siendo objeto de duras críticas, la Lingüística de Corpus experimentó una época difícil. Fue empujada al margen de la corriente lingüística dominante. Sólo unos pocos lingüistas continuaron con la recogida y la inducción de datos. Gracias a su trabajo, no sólo reformaron el método y las herramientas de compilación y tratamiento de recursos de corpus, sino que también ampliaron el campo de investigación basado en corpus.

En 1959, Quirk planteó el proyecto SEU (*Survey of English Usage*), cuyo objetivo era hacer una descripción completa y adecuada de la gramática del inglés. Al principio, los datos también fueron copiados en fichas. Poco después, en 1961, el *Brown Corpus*, dirigido por N. Francis, H. Kucera, entre otros lingüistas y expertos informáticos, fue el primer corpus informatizado en el mundo. A partir de ese momento, nació la primera generación de corpus modernos, marcada por la presencia del ordenador.

La primera generación de corpus informatizados contaba con las siguientes características:

**Formato electrónico:** En comparación con los corpus tradicionales, este formato facilitaba a los usuarios el almacenamiento, la búsqueda, el procesamiento o el mutuo compartir de recursos.

**Representatividad de los datos:** Los proyectos de elaboración de corpus pretendían recoger textos que representaran el estado de la lengua en ese momento. Se produjeron más corpus de textos escritos que orales.

**Tamaño:** Como el ordenador podía almacenar grandes recursos, la velocidad de procesamiento y la magnitud de almacenamiento aumentaron mucho. Ciertos corpus llegaban a contar cerca del millón de palabras.

Algunos ejemplos de esta primera generación de corpus son los siguientes:

- *The Brown Corpus*. Se compiló en la Universidad de Brown, con 500 muestras de 2000 palabras (1 millón de palabras) del inglés americano escrito.
- *Lancaster-Oslo/Bergen Corpus* (LOB). Se compiló en Lancaster, Oslo y Bergen en 1961, con 1 millón de palabras de inglés británico escrito.
- *The London-Lund Corpus of Spoken English* (LLC). Derivaba de dos proyectos. El primero fue *The Survey of English Usage Corpus* (SEU) de la *University College London*, lanzado en 1959 por Randolph Quirk. El segundo era *The Survey of Spoken English* (SSE), dirigido por Jan Svartvik en la *Lund University* en 1975. Este corpus despertó el interés de la investigación por el inglés hablado.
- *The Lancaster/IBM Spoken English Corpus* (SEC). Contenía 52.000 palabras del inglés británico oral. Como su etiquetador abarcaba la anotación fonética y gramatical, se usaba en la investigación de síntesis de habla y reconocimiento fonético.

En la década de los 80, los corpus eran capaces ya de mostrar muchas ventajas. Se utilizaban en el estudio del lenguaje como test de hipótesis lingüísticas, construcción de sistemas de procesamiento del lenguaje natural, etc. Dado su gran tamaño, sus diversas funciones, su gran capacidad de procesamiento y almacenamiento y facilidades de búsqueda, la investigación basada en corpus electrónicos cada vez fue llamando más la atención de los investigadores. Diversos lingüistas disputaban y argumentaban a favor del papel de los corpus en la investigación lingüística. Esta metodología basada en recursos de lengua natural volvió, así, a lograr el reconocimiento general.

Los grandes corpus de esta nueva generación se caracterizan por:

**Rapidez:** El desarrollo de internet y la velocidad de descarga hacen que los datos y resultados se puedan transferir de lingüista a lingüista de manera muy rápida.

**Tamaño:** Con el desarrollo tecnológico y el progreso en la elaboración y empleo de programas específicos para la lingüística de corpus, cada vez aparecen corpus más grandes, sobrepasando los mil millones de palabras.

**Internacionalización:** Empiezan a prestar atención a cada variante de la lengua. La creación de corpus de inglés se inició en Inglaterra y los Estados Unidos. A medida que el inglés se popularizaba en el mundo, los lingüistas se daban cuenta de que también eran importantes los corpus que pudieran representar el inglés como lengua segunda o lengua extranjera. Por ejemplo, el corpus *The International Corpus of English* (ICE) tenía por objeto la colección de recursos orales y escritos para hacer comparación del estudio de inglés en el mundo. 23 equipos de investigación a nivel mundial han participado en este proyecto. Cada equipo seguía un diseño común y un programa de anotación gramatical común<sup>1</sup>.

**La aparición de corpus monitor dinámicos, actualizados continuamente:** Por ejemplo, cada semana, el corpus COBUILD de Inglaterra, envía *e-mails* a sus usuarios, mandándoles información sobre cambio dinámico de palabras y estructuras.

**Lenguas:** En la fase inicial del desarrollo de la Lingüística de Corpus, los corpus en la lengua inglesa ocupaba la posición dominante. Poco a poco, empiezan a florecer corpus en diferentes lenguas y corpus multilingües.

**Procesamiento:** La automatización del procesamiento de los textos de un corpus (asignación de categoría gramatical, desambiguación, extracción de concordancias, etc.) es cada vez más potente.

**Facilidad de acceso.** Muchos recursos de corpus están disponibles en Internet. Como consecuencia, no hacen falta herramientas complicadas para acceder a ellos.

Como ejemplos de corpus de esta segunda generación, destacamos:

---

<sup>1</sup> *The International Corpus of English* [en línea]: < <http://ice-corpora.net/ice/index.htm> > [consulta: 13 de abril de 2011]

- *Longman Corpus Network*. Formado por tres corpus que son *Longman/Lancaster English Language Corpus* (LLELC), *Longman Spoken Corpus* (LSC) y *Longman Corpus of Learners' English* (LCLE). Uno de los objetivos es la elaboración de diccionarios para los aprendices extranjeros del inglés. Consta de 50 millones de palabras.
- *The British National Corpus* (BNC). Contiene 100 millones de palabras del inglés británico contemporáneo escrito y hablado.
- *The Bank of English* (Proyecto COBUILD). Contiene 524 millones de palabras del inglés moderno escrito y oral.
- *Corpus de Referencia del Español Actual* (CREA). Contiene 160 millones de palabras de textos escritos y orales del español actual.
- *Corpus Diacrónico del Español* (CORDE). Abarca textos escritos del español desde la Edad Media hasta la época contemporánea, con 250 millones de palabras.
- *Corpus for Spoken Romance Languages* (C-ORAL-ROM). Se trata de un corpus oral multilingüe y general del habla de cuatro lenguas romances (francés, español, italiano y portugués) en sus diferentes registros.
- *Global Language Monitor* (GML). Se estableció como proyecto en 2003. Su principal objetivo es registrar, analizar y seguir la tendencia de desarrollo de las lenguas en el mundo, sobre todo, el desarrollo del inglés. El creador Paul Payack planteó el índice PQI (*The Predictive Quantities Indicator*) para seguir la frecuencia del vocabulario inglés en los medios de comunicación y predecir tendencias en su desarrollo, así como su cambio a corto plazo, etc.

### **2.1.4 Tipos de corpus**

Existe una gran variedad de tipos de corpus, que difieren entre sí conforme a múltiples aspectos. En este apartado, se enuncian algunos de los principales tipos de corpus según el propósito y los parámetros.

Según la modalidad de la lengua,

**Corpus escritos:** también llamados corpus textuales, son muestras de lengua escrita.

**Corpus orales:** muestras de lengua hablada. Se pueden dividir en (Moreno y Urresti, 2005):

- **Corpus de habla espontánea:** transcripciones de producciones lingüísticas espontáneas.
- **Base de datos acústicos:** el núcleo de la investigación es la señal sonora, y se caracterizan por su carácter controlado.

**Corpus mixtos:** combinación de muestras de lengua escrita y lengua hablada.

**Corpus multimodal:** muestras lingüísticas en varios formatos multimedia (texto, sonido, imagen, etc.). Gracias a ellos, se puede leer la transcripción, escuchar la grabación y visualizar la imagen a la vez. Es un nuevo tipo de corpus.

Según el número de lenguas,

**Corpus monolingües:** muestras de una sola lengua. Se recopilan con el objetivo de dar a conocer una lengua o variedad lingüística.

**Corpus bilingües o multilingües:** muestras de dos (bilingües) o más lenguas (multilingües). Son útiles para un aprendizaje constructivo a partir del conocimiento previo de la propia lengua, así como para la detección de “falsos amigos”. Se pueden dividir en:

- **Corpus comparables:** textos parecidos respecto a sus características y criterios de selección. Se pueden usar para la comparación de diferentes lenguas. C-ORAL-ROM pertenece a este tipo.



- **Corpus paralelos:** el mismo texto original traducido a una o más lenguas representadas en el corpus. Se pueden utilizar en la traducción automática o en entornos bilingües o multilingües como la ONU, la UE, la OTAN, etc.

Según el procesamiento,

**Corpus puros:** textos electrónicos guardados sin ninguna anotación.

**Corpus anotados:** textos enriquecidos mediante la adición de información lingüística de algún tipo (ortográfica, fonética, fonológica, prosódica, semántica, discursiva o gramatical: morfosintáctica o sintáctica) (Santalla del Río, 2005).

**Corpus verticales:** el resultado de disponer en forma de columna las palabras de un texto ordenadas según criterios alfabéticos o de frecuencia. Las palabras se tratan así aisladamente, sin contexto (Villayandre, [en línea]).

Según la especificidad de los textos,

**Corpus generales:** también llamados corpus de referencia; muestras que tratan de reflejar la lengua y sus variedades lingüísticas de la manera más exhaustiva y equilibrada posible.

**Corpus especializados:** textos compilados para la descripción de uno o varios aspectos concretos de lengua.

Según criterios de temporalidad,

**Corpus cronológicos:** muestras de uno o más períodos concretos con el objeto de estudiar la lengua producida durante esa horquilla de tiempo.

**Corpus diacrónicos:** muestras de diferentes etapas para el estudio de la evolución de la lengua durante un largo período.

**Corpus sincrónicos:** muestras de una o más variedades lingüísticas en un momento histórico dado. Se pueden utilizar para hacer comparaciones entre variedades o dialectos de una lengua.

### ***2.1.5 La aplicación de corpus***

Después de entrar en los años 80, el desarrollo de la Lingüística de Corpus se ha acelerado. Se han compilado muchos corpus nuevos. El alcance de su aplicación se amplía gradualmente:

**Redacción de libros de referencia:** Como colecciones extensas de textos naturales, los corpus constituyen una fuente muy rica de materiales. Pueden suministrar gran cantidad de ejemplos y reflejar el uso más típico y nuclear de la lengua contemporánea. Muchos diccionarios o gramáticas se redactan sobre la base de corpus. Por ejemplo, después del establecimiento del corpus SEU, Quick y sus colegas confeccionaron *A Grammar of Contemporary English* y *A Comprehensive Grammar of the English Language*. Basándose en el proyecto COBUILD, se publicaron los siguientes libros: *The Collins COBUILD English Language Dictionary*, *The Collins COBUILD English Grammar*, *The Collins COBUILD Student's Grammar*, *The Collins COBUILD Dictionary of Phrasal Verbs*, *Collins COBUILD English Guides*, *Collins COBUILD English Usage*, etc.

**Investigación de la lengua:** Con muestras auténticas y abundantes, los corpus repercuten muy positivamente en la investigación de la lengua. Su creación ha reformado la metodología de la investigación teórica de la lengua. Los corpus diacrónicos que abarcan una serie de etapas históricas son útiles para el estudio de la historia de la lengua. La función de herramientas como los programas de concordancias y el recuento de palabras pueden facilitar mucho la descripción de las lenguas y el análisis lingüístico (sintáctico, fonético, discursivo, gramatical, semántico...). Ayudan a la comparación entre variedades regionales o entre la lengua hablada y la escrita. Gracias a los corpus, las teorías lingüísticas disponen de ejemplos reales y datos cuantitativos.

**Lingüística Computacional:** La Lingüística Computacional trata de la construcción de sistemas informáticos que procesen la estructura lingüística y cuyo objetivo sea la simulación parcial de la capacidad lingüística de los hablantes de una lengua (Moreno, 1998). Con los datos extraídos de los corpus, se ha impulsado el desarrollo de la Lingüística Computacional, aplicada a tareas como la traducción automática, la investigación de herramientas de procesamiento de corpus, el reconocimiento de habla, etc. Ambos son complementarios. Se puede lograr, sin lugar a dudas, una interacción fructífera.

Con el desarrollo de corpus, nacieron una serie de herramientas automáticas de anotación como analizadores sintácticos, etiquetadores morfológicos, etc. Tienen capacidad de tratar grandes cantidades de muestras. A pesar de que el resultado no puede llegar a ser completamente correcto, son muy prácticas en la elaboración y la anotación de corpus. Efectivamente, antes la anotación se realizaba manualmente. Ahora, la combinación de procesamiento automático y corrección manual han acelerado en gran medida la velocidad de anotación. En este sentido, la tecnología de la Lingüística Computacional sirve como la base técnica de la creación y procesamiento de corpus.

Por otra parte, los investigadores de la Lingüística Computacional pueden acceder fácilmente a corpus de lengua escrita y oral de gran tamaño, entre ellos se incluyen corpus anotados como *Penn Treebank*, *Prague Dependency Tree Bank*, *PropBank*, *RST Bank*, *Time Bank*, etc. Son recursos que muestran anotación de diferentes estratos como la sintáctica, la semántica, la retórica, etc. Su existencia hace que la investigación de la Lingüística Computacional pueda usar la metodología de aprendizaje automático supervisado (*supervised machine learning*) para tratar aquellos análisis automáticos tradicionalmente complicados como el análisis sintáctico o semántico (Feng Zhiwei, 2008).

La traducción automática atañe a la tecnología de procesamiento de la lengua natural y de diferentes lenguas. Hay que reconocer que todavía faltan herramientas de análisis perfeccionadas para la traducción de una oración o un

párrafo completo. Pero los traductores humanos pueden introducir la lengua original y corregir luego manualmente, lo cual ahorra mucho tiempo de la traducción. Los corpus monolingües pueden servir para la traducción automática indirectamente. Los corpus paralelos pueden servir para la traducción automática directamente. Siguiendo determinados criterios y requisitos, se están desarrollando muchos programas sobre la base de la investigación de corpus grandes. El informático de la USC, Franz Och, incluso llegó a decir “Give me enough parallel data, and you can have a translation system for any two languages in a matter of hours (Mackin, [en línea])”.

**Enseñanza de lenguas:** Los corpus contienen gran cantidad de datos de lengua de uso real, que pueden usarse en la enseñanza de lenguas maternas y lenguas extranjeras. Aprovechándolos, se puede preparar el vocabulario según la frecuencia de palabras; se puede ayudar a los alumnos a entender el significado y uso de algunas palabras en la situación comunicativa real; se puede reducir la diferencia entre la lengua enseñada en clase y la lengua usada por los nativos; se puede evaluar el nivel de la lengua; se puede facilitar el estudio fuera de clase; se pueden analizar más eficazmente los errores. En el Capítulo 3, se explica con mayor detalle esta polivalente aplicación de corpus.

### ***2.1.6 Ventajas de los corpus***

La creación de corpus, unida al desarrollo de herramientas informáticas, ha hecho realidad la posibilidad de un análisis efectivo de grandes colecciones de datos. Hasta hoy, los corpus siguen teniendo sus limitaciones. Es imposible conseguir la representatividad absoluta. En el caso de corpus de lengua oral, todavía cuesta mucho tiempo hacer las transcripciones. Aunque los corpus monitor tienen actualización continua, muchos corpus están acotados temporalmente. Las colecciones de muestras escritas u orales tienen fecha inicial y final. Para una época determinada, pueden ser representativos; sin embargo, más adelante, pueden resultar anacrónicos. Las herramientas no son totalmente automáticas y se pueden mejorar. Sin embargo, son muy obvias sus ventajas:

- Textos en formato electrónico. Las herramientas informáticas se pueden aplicar con gran rapidez en el procesamiento de operaciones tales como la búsqueda, la recuperación de información o el cálculo de frecuencias. Es evidente que ahorran mucho tiempo y mano de obra.
- El corpus permite la investigación de muestras reales y de datos estadísticos, lo cual ha elevado la fiabilidad y la autenticidad de los estudios lingüísticos.
- Reutilización de recursos. Un mismo corpus puede usarse para diferentes investigaciones y en diferentes disciplinas científicas. Además, los corpus tienen un carácter abierto: se pueden componer nuevos corpus aprovechando corpus existentes y añadir nuevos contenidos.
- Son un recurso necesario para la exploración de los aspectos cuantitativos del lenguaje.
- Contiene nuevos fenómenos lingüísticos. Podemos encontrar en un corpus ciertas palabras nuevas que los diccionarios no contemplan y construcciones sintácticas que no aparecen en las gramáticas.

A modo de resumen, gracias al desarrollo tecnológico y la profundización de la investigación, los corpus son cada vez más completos y más imprescindibles.

## ***2.2 El corpus C-ORAL-ROM***

Según Massimo Moneglia, el habla espontánea se puede definir como “every oral performance which does not execute a previous (written or scripted) text (Cresti y Moneglia, 2005: 4)”. Según esto, los eventos de habla espontánea comprenden:

- Interacción multimodal cara a cara;
- Referencia intersubjetiva a un espacio deíctico;
- Programación mental simultánea con la ejecución vocal (sin guión);

- Comportamiento lingüístico contextualmente indeterminado (comportamiento impredecible) (Cresti y Moneglia, 2005).

En comparación con corpus de lengua escrita, los corpus orales florecen más tarde que los corpus escritos. Debido a las dificultades de la transcripción, los corpus orales suelen ser mucho más pequeños en tamaño.

En las últimas dos décadas, se han notado avances importantes en la recopilación y desarrollo de corpus orales, sobre todo en el idioma inglés. Algunos ejemplos son:

- *The Cambridge and Nottingham Corpus of Discourse in English;*
- *The Limerick Corpus of Irish English;*
- *The Hong Kong Corpus of Spoken English;*
- *The Michigan Corpus of Academic Spoken English.*

No cabe duda de su importancia. Los corpus orales pueden proporcionar un recurso único para el análisis de discursos naturales, y el creciente interés en el desarrollo de corpus orales sirve como prueba del valor que ofrecen a la comunidad científica (Adolphs y Knight, 2010).

En este apartado, nuestra intención es resaltar especialmente el corpus C-ORAL-ROM. En la elaboración de C-ORAL-CHINA, nos ha servido como referencia y seguimos muchas de sus convenciones en la transcripción.

C-ORAL-ROM es un corpus oral multilingüe de lengua espontánea en las cuatro lenguas romance principales: francés, italiano, portugués y español. Se compone de 772 textos y aproximadamente 121 horas espontáneas de habla por parte de 1,427 hablantes. Cada lengua tiene alrededor de 300.000 palabras (Cresti y Moneglia, 2005). Este proyecto fue financiado por la UE bajo el *V Framework Programme* (IST-2000-26228) y el consorcio está formado por nueve socios, coordinados por la Universidad de Florencia. El principal objetivo del proyecto

C-ORAL-ROM es “to provide a comparable set of corpora of spontaneous spoken language of the main Romance languages (Cresti y Moneglia, 2005: 1)”.

La característica más significativa de C-ORAL-ROM consiste en su espontaneidad: los textos se han grabado en su contexto real y sin guión alguno. Cada subcorpus cuenta con la misma distribución textual para asegurar la comparabilidad y la representatividad. El recurso se presenta en varios formatos: una transcripción ortográfica, una versión etiquetada en XML y el alineamiento de sonido y texto. Además, se proporcionan una anotación lingüística parcial de los textos y programas para manejar los recursos y los estudios cuantitativos.

Massimo Moneglia (2005) explica las estrategias del diseño de C-ORAL-ROM del modo siguiente:

- Se divide en habla informal (50%) y habla formal (50%).
- En las dos partes, se adoptan distintos criterios de selección de muestras.
- Define la dimensión de muestras en términos de número de palabras.
- Asegura la representación suficiente del habla informal dialogal, que tiene mayor valor.
- Garantiza la representación de habla formal tanto en medios de comunicación como en contextos naturales (*face to face*).

La Tabla 2.1<sup>2</sup> nos muestra la matriz del diseño de C-ORAL-ROM.

---

<sup>2</sup> Laboratorio de Lingüística Informática [en línea]: < <http://www.llif.uam.es/ESP/> > [consulta: 16 de enero de 2011]

Tabla 2.1 Matriz del diseño de C-ORAL-ROM

			<b>Requisitos</b>	<b>Italiano</b>	<b>Francés</b>	<b>Español</b>	<b>Portugués</b>
<b>Sección</b>	<b>Contexto</b>	<b>Dominio</b>	<b>Palabras</b>	<b>Palabras</b>	<b>Palabras</b>	<b>Palabras</b>	<b>Palabras</b>
<b>INFORMAL</b>			150000	155048	140341	169625	167059
	<b>Familiar-privado</b>		124500	128696	113287	131734	134511
		<b>Monólogo</b>	42000	45213	42436	42718	45939
		<b>Diálogo-Conversación</b>	82500	83483	70851	89016	88572
	<b>Público</b>		25500	26352	27054	37891	32548
		<b>Monólogo</b>	6000	6051	6521	6182	7693
		<b>Diálogo-Conversación</b>	19500	20301	20533	31709	24855
<b>FORMAL</b>			150000	156544	129618	165846	152483
	<b>Contexto natural</b>		65000	68324	47862	72573	66151
	<b>Medios de comunicación</b>		60000	61638	54176	62809	62018
	<b>Teléfono</b>		25000	26582	27580	30464	24314
<b>TOTAL</b>			300000	311592	269959	335471	319542



Según sus estrategias, la longitud del texto se determina en términos de cantidad de información (por ejemplo, número de palabras). Las diferentes estrategias adoptadas en la sección formal e informal son:

En la sección informal:

- **Textos cortos:** al menos 64 textos de unas 1.500 palabras cada uno. Hasta un 20% de esta parte puede estar constituida por textos de diferentes longitudes.
- **Textos largos:** de 8 a 10 textos de unas 4.500 palabras cada uno.

En la sección formal, la longitud del texto se define según las siguientes reglas:

- **Para Formal en contexto natural:** 2 ó 3 muestras para cada dominio de una media de 3.000 palabras.
- **Para Medios de comunicación:** 2 ó 3 muestras por cada dominio. Una media de 3.000 palabras. Sólo muestras breves para Partes meteorológicas y Noticias.
- **Para Teléfono:** longitud del texto no definida (preferentemente 1.500 palabras de límite máximo, sin límite mínimo). El dominio de las llamadas a servicios telefónicos y/o interacciones humano-máquina debe contener 10.000 palabras. Este campo no está representado en el corpus portugués.

En cuanto a la calidad de grabación, cabe señalar que el proyecto C-ORAL-ROM está orientado a la recogida de datos para un corpus en un ambiente natural, a pesar del hecho de que esto provoca necesariamente una disminución de la calidad acústica del recurso. Aparte, C-ORAL-ROM ha explotado, en el marco de un nuevo trabajo multilingüe, la riqueza de contenidos que suponen los archivos recopilados durante años de investigación en lengua oral.

Por eso, la calidad acústica y las condiciones de grabación de las fuentes son variables. En este aspecto nuestro corpus difiere, ya que C-ORAL-CHINA es un corpus eminentemente destinado a la enseñanza y exige una alta calidad acústica.

En C-ORAL-ROM, los requisitos para el formato acústico y los aparatos de grabación son los siguientes:

- **Formato:** archivos .wav mono (Windows PCM), frecuencia de las muestras: 22050Hz, 16 bit.
- **Proceso de grabación y almacenamiento para antiguas grabaciones analógicas:** se convierten directamente en archivos wav (20.050Hz 16 bit) mediante una tarjeta de sonido estándar (Sound Blaster live o compatible) con un editor de sonido profesional, Cool Edit 2000 ®.
- **Proceso de grabación y almacenamiento para nuevas grabaciones:**
  1. Diálogos: grabación estéreo con DAT o minidisk (44.100Hz) con micrófonos unidireccionales, convertidos en archivos wav mono (Windows PCM, 22050Hz, 16 bit) por medio de un puerto SPDIF de una tarjeta de sonido estándar (Sound Blaster live o compatible) con un editor de sonido profesional.
  2. Conversaciones con más de dos participantes: grabación mono con DAT o minidisk con micrófono omni-direccional convertida en archivos wav mono mediante un puerto SPDIF de una tarjeta de sonido estándar (Sound Blaster live o compatible) con un editor de sonido profesional.

Los archivos de habla de la base de datos acústica se definen en una escala de calidad (grabación, volumen, solapamiento de voz y ruido). La escala de calidad se extiende desde el nivel más alto de claridad de la señal de voz hasta los niveles más bajos de calidad acústica.

1. Grabaciones digitales con DAT o minidisk y micrófonos unidireccionales o grabaciones analógicas de alta calidad.
2. Grabaciones digitales con respuesta de micrófono más pobre o grabaciones analógicas con:
  - Buena respuesta de micrófono
  - Bajo ruido de fondo
  - Bajo porcentaje de enunciados solapados
  - Cálculo de la F0 posible en la mayoría de los archivos
3. Grabaciones analógicas de baja calidad con:
  - Respuesta de micrófono pobre
  - Ruido de fondo
  - Porcentaje medio de enunciados solapados
  - Cálculo de la F0 posible en muchas partes de los archivos

La calidad se evalúa espectrográficamente. Las secciones en las que el análisis de la F0 no es significativo se excluyen del muestreo. La calidad acústica de cada grabación y los datos más relevantes sobre las condiciones de grabación se incluyen siempre en los metadatos de cada texto.

Los archivos de habla y de transcripción están en correspondencia de uno a uno. A continuación se presenta la tabla general del corpus multimedia C-ORAL-ROM:

	<b>archivos .wav</b>	<b>GB</b>	<b>Duración</b>	<b>Enunciados</b>	<b>Palabras en archivos .txt</b>
<b>Francés</b>	206	3,77	26.21.43	19546	256271
<b>Italiano</b>	204	5,19	36.16.10	40402	311592
<b>Portugués</b>	152	4,43	29.43.42	38855	317920
<b>Español</b>	210	4,56	31.06.00	35588	335471

Tabla 2.2 Tabla general del corpus multimedia C-ORAL-ROM  
(Cresti y Moneglia, 2005)

Para cada sesión de grabación de habla espontánea, se proporcionan los siguientes archivos en las carpetas del corpus multimedia.

1. Archivos de habla: archivos .WAV descomprimidos (Windows PCM: 22,050 hz; 16 bit).
2. Transcripciones en formato CHAT, enriquecidas con la anotación de rupturas prosódicas terminales y no terminales y la información sobre el alineamiento, en archivos TXT.
3. Archivos de alineamiento sonido-texto: archivos XML en formato WIN PITCH CORPUS.
4. DTD del formato de alineamiento WinPitchCorpus.
5. Transcripción de cada sesión en formato CHAT en archivos .TXT (sin la información de alineamiento).
6. Transcripción de C-ORAL-ROM para cada sesión en archivos .XML.
7. DTD para el formato C-ORAL-ROM .XML.
8. Metadatos en formato CHAT.
9. Metadatos en formato IMDI.
10. Transcripción de C-ORAL-ROM de cada sesión con anotación de categorías y lemas para cada forma en archivos .TXT.
11. Etiquetario adoptado en archivos .TXT.
12. Listas de frecuencias de lemas y formas en archivos .TXT.
13. Medidas de los valores lingüísticos grabados en cada texto: en los archivos Excel "measurements\_language.xls".
14. Diagramas que presentan la tendencia observada en relación a los parámetros de variación textual estándar a lo largo de los nodos estructurales del corpus, en el archivo Excel "Multilingual graphics.xls".

### 2.3. La Lingüística de Corpus en China

El libro 语体文应用字汇 fue publicado por *The Commercial Press* en 1928. Al editarlo utilizando el método de contabilidad, el pedagogo Chen Heqin creó un corpus de la lengua escrita con el objetivo de calcular la frecuencia de caracteres chinos e instituir una tabla de caracteres básicos de chino. Esta experiencia fue considerada como el inicio de la investigación de corpus en China (Chen Heqin, 1928).

A partir de los años 70, la Lingüística de Corpus en China ha estado en ascenso constante.

Los primeros corpus informatizados son:

- Corpus de Obras Literarias en Chino Moderno (1979). Contiene 5.270.000 de caracteres. Se compiló por la Universidad Wu Han.
- Corpus de Chino Moderno (1983). Contiene 20.000.000 de caracteres. Se compiló por la Universidad de Aviación de Beijing.
- Corpus de Manuales de Chino de Escuela Secundaria (1983). Contiene 1.068.000 caracteres. Se compiló en la Universidad Pedagógica de Beijing.

Estos corpus se caracterizan por:

1. Como muchos fueron creados manualmente, se necesitaba mucha gente y costaba mucho tiempo. El tamaño todavía era pequeño.
2. Se estudiaban la frecuencia de palabras, la segmentación automática de palabras y la segmentación ambigua.
3. Establecimiento del estándar de la segmentación de palabras. Como es sabido, en la escritura, el idioma chino no tiene espacio entre palabra y palabra. En 1990, se elaboró un estándar nacional llamado GB-13715 como la base de la segmentación automática de lengua escrita de chino.

A partir de los años 90, cada vez aparecen más corpus de gran tamaño.

Ejemplos:

- Corpus de Diario del Pueblo. Colecciona todos los artículos e imágenes de este diario durante 48 años.
- Corpus Oral de Beijing Contemporáneo. En los años de 80, la Universidad de Lengua y Cultura de Beijing hizo la grabación del lenguaje hablado de 374 pequineses y la transcripción de 1.860.000 caracteres<sup>3</sup>.
- Corpus Estatal de Chino Moderno. Se inició en 1993 por el Comité Estatal de Lenguas. Los primeros datos abarcaban muestras de la lengua escrita desde 1919 hasta 1992. En total, son 70.000.000 caracteres. Después, cada año se añaden del orden de 10.000.000 caracteres.
- Corpus de Estudio de Gramática de Chino Moderno. Fue compilado en 1995 por la Universidad de Lengua y Cultura, con 20.000.000 caracteres. Todo el corpus cuenta con segmentación de palabras y anotación de categorías.
- Corpus de Chino Moderno. Fue compilado en 1998 por la Universidad Qing Hua con 10.000.000 caracteres. Su objetivo era estudiar la segmentación ambigua.
- Corpus LIVAC, *Linguistic Variety in Chinese Communities*. Se recolecionó simultáneamente el chino que se usa en cinco zonas de habla china (parte continental, Hong Kong, Macao, Taiwan, Singapur) para hacer una descripción y análisis completo del uso de chino en diferentes lugares. Hasta junio de 1996, los caracteres sacados desde la prensa llegaron a 830.000.000.

Características:

- Cada vez el tamaño era mayor.

---

<sup>3</sup> “《当代北京口语语料库》说明”: [en línea]:

< [http://www.blcu.edu.cn/yys/6\\_beijing/yuliao/shuoming.htm](http://www.blcu.edu.cn/yys/6_beijing/yuliao/shuoming.htm) > [consulta: 17 de abril de 2011]

- El procesamiento fue siendo cada vez más profundo: han aparecido diferentes métodos de segmentación de palabras, una serie de anotación automática de categorías, anotación de la estructura de locución que puede expresarse en *P-Tree* y *D-Tree*, sistema de análisis morfológico, sistema de análisis sintáctico, etc.
- Gran cobertura: incluye ciencias sociales y naturales.
- Método de la combinación del trabajo automático y revisión personal.
- Derecho intelectual: cada vez supone un elemento a tener más en cuenta.
- Tipología de corpus: abarcan casi todos los tipos de corpus. Por ejemplo, para el corpus bilingüe o multilingüe, aparecieron corpus de chino e inglés, corpus de japonés y chino, corpus de alemán y chino, corpus de chino, japonés e inglés, etc. Para el corpus de inglés, se compilaron corpus paralelos, corpus de inglés para uso especial, corpus de aprendices de inglés (lengua escrita y hablada), etc.
- La fundación de consorcios. Dos consorcio famosos son:

*Chinese Linguistic Data Consortium* (ChineseLDC, 中文语言资源联盟, <http://www.chineseldc.org/>)

*Chinese Corpus Consortium* (CCC, 国际中文语言资源联盟, <http://www.d-ear.com/ccc/index.htm>)

Tomamos el propósito de CCC<sup>4</sup> como ejemplo:

- Compilar e integrar recursos lingüísticos en chino ya existentes, y continuar creando nuevos recursos.
- Integrar las herramientas existentes para la creación, la transcripción y el análisis del habla espontánea en chino, mejorando su utilización, y crear nuevas herramientas.
- Crear especificaciones y normas para definir los estándares

---

<sup>4</sup> Tomado de < <http://www.d-ear.com/ccc/about.htm> > [consulta: 13 de marzo de 2011]

de las tecnologías lingüísticas en chino.

- Promover el intercambio de recursos lingüísticos en chino.

Se ve que su objetivo es compartir los recursos, promover la investigación conjunta y posibilitar intercambios.



### 3. El uso de corpus en la enseñanza de lenguas

#### *3.1 Las aplicaciones de corpus en la enseñanza de lenguas*

En el apartado 2.1.5 se resumen las aplicaciones de corpus en diferentes áreas entre ellas, en la enseñanza de lenguas, donde el uso de corpus y sus herramientas de análisis se han incrementado en las últimas dos décadas. Tanto profesores como alumnos pueden usarlos en clase y fuera de ella. En este apartado, se abordan las aplicaciones de corpus especialmente para la enseñanza de lengua segunda.

El valor de los corpus en este campo ha ido reconociéndose cada vez más entre los estudiosos y profesores. Muchos investigadores han discutido sobre ello y han expuesto sus criterios al respecto. Sinclair (2004) propugna la enseñanza de lenguas basada en corpus (*corpus-based language teaching*) como una nueva revolución en la enseñanza de lenguas. Fligelstone (1993) describió tres propósitos del uso de corpus en la enseñanza: “teaching about (the principles and theory behind the use of corpora), teaching to exploit (the practical, methodological aspects of corpus-based work), and exploiting to teach (using corpora to derive or drive teaching materials)”. Renouf (1997) añadió el cuarto “teaching to establish resources” (田omado de Cheng, 2010: 319). McEnery (2009: 4) es de la opinión de que “the use of language corpora in the classroom is slowly being tested and established as an indispensable tool. This is particularly true in the area of second language pedagogy, where multilingual and learner corpora are stimulating much in the way of innovation in the classroom”.

Aparte, se celebran conferencias internacionales como *Teaching and Language Corpora* (TaLC) para que los investigadores puedan intercambiar experiencias y profundizar en la investigación de las aplicaciones.

Leech (1997) propuso tres enfoques de las aplicaciones de uso de corpus en la didáctica de lenguas: el uso directo, el uso indirecto (publicaciones, materiales, test, etc.) y el desarrollo de corpus orientados a la enseñanza (*LSP corpora*, *L2 learner corpora*, etc). La Figura 3.1 presenta la estructura de la investigación en función de estos tres enfoques.

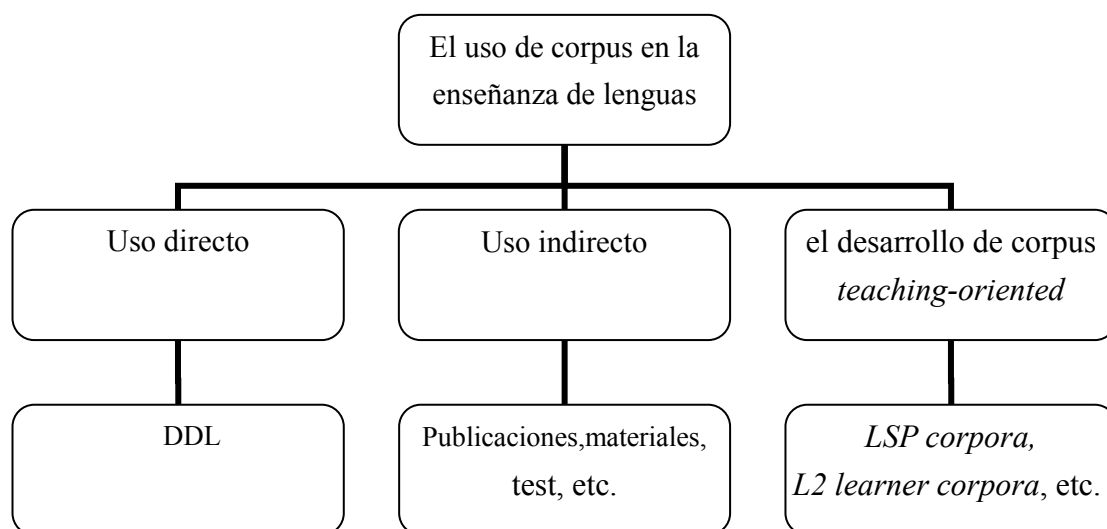


Figura 3.1 Aplicaciones de Corpus en la Enseñanza de Lenguas

### 3.1.1 DDL

El primer enfoque influye directamente en el proceso de la enseñanza. Johns y King (1991: 111) fueron pioneros en introducir el concepto DDL (*Data-driven Learning*) y lo definieron como “Data-driven learning is the use in the classroom of computer generated concordances to get students to explore regularities of patterning in the target language, and the development of activities and exercises based on concordance output”. Johns identificó tres etapas de carácter inductivo en el uso de corpus según el método DDL: “observation (of concordanced evidence), classification (of salient features) and generalization (of rules)”.

Tradicionalmente, la metodología en la enseñanza de lenguas se adecua al esquema “Three Ps (presentation-practice-production)”. Bajo esta nueva metodología, sin embargo, los alumnos de segunda lengua cuentan con más autonomía. En efecto, según este nuevo modelo conocido como “Three Is (illustration-interaction-induction) (Carter y McCarthy, 1995: 155)”, los alumnos pueden observar y analizar los ejemplos reales e inducir reglas lingüísticas mediante la consulta directa de gran cantidad de datos lingüísticos que ofrecen los corpus.

Esta metodología alienta a los alumnos a estudiar a través del propio “descubrir”. Trata de despertar el interés por el estudio en los alumnos, procura formarles en la capacidad de resolver problemas y pretende mejorar la intuición de los alumnos.

### ***3.1.2 El uso de corpus en la elaboración de manuales y diseño de programas de estudio***

Una característica incomparablemente ventajosa de los corpus estriba en la riqueza de sus recursos, llenos de realismo lingüístico. Braun (2005) afirmó que al describir una lengua basándose en los corpus se obtenían datos realistas, ricos, ilustrativos y actualizados que pueden utilizarse como un recurso para la creación de un material didáctico interesante. Gracias a esto, pueden favorecer de manera indirecta la enseñanza de lenguas, en aspectos tales como el desarrollo de materiales, el diseño de programas de estudio, la evaluación de nivel de lengua, etc.

En el apartado 2.1, ya hemos dado el ejemplo del uso del corpus COBUILD para la redacción de diccionarios y gramáticas.

La elaboración de manuales convenientes y el diseño de programas de estudio son dos elementos muy importantes para la enseñanza de lenguas. Sin embargo, en muchos casos, se toman la experiencia o la

intuición de los autores como base. Como no se aplica el estándar de índices cuantitativos objetivos, a veces, no pueden garantizar que se adapten a la necesidad de la enseñanza y provocan, como consecuencia, una distancia distorsionadora entre aquello que se enseña en el extranjero y lo que usan los nativos en la vida real.

Por ejemplo, en el estudio de actos de habla para expresar desacuerdo, Cheng y Warren (2005) han apuntado que los autores de manuales de inglés deberían incorporar más formas en sus materiales con el objetivo de reflejar mejor la realidad del uso de la lengua actual. Römer (2005) hizo una comparación de las funciones y los patrones contextuales de las formas progresivas del inglés entre el material que se emplea en los colegios alemanes en la enseñanza del inglés como lengua extranjera y el inglés usado en BNC (*British National Corpus*). Descubrió una distorsión entre ambos. Señaló que las formas gramaticales “progresivas” que se refieren a acciones o eventos repetidos, eran mucho más frecuentes en el inglés real que en el inglés de los libros de texto, en los que la función común de repetición estaba bastante abandonada y se centraba en eventos continuos.

Los corpus pueden servir como base para los manuales o los programas de estudio en muchos aspectos como los listados de vocabulario, la identificación de estructuras gramaticales recurrentes, el orden de su aparición, etc. Se pueden utilizar los listados de frecuencia de palabras o las estructuras gramaticales derivadas del uso de corpus como punto de partida para la confección de los listados del vocabulario y los puntos gramaticales que se deben enseñar en cada nivel de un manual. Además, por medio del análisis de la distribución y la frecuencia, la herramienta de concordancias nos permite juzgar el nivel, más o menos elevado, de aprendizaje de una lengua que propone cada manual.

### ***3.1.3 El uso de corpus en exámenes de lenguas***

Los exámenes forman parte imprescindible de la enseñanza de lenguas. Cómo preparar un examen de buena calidad es un trabajo complicado. En 1996, Alderson propuso la potencialidad de la aplicación de corpus en diferentes etapas de exámenes de lenguas (Alderson, 1996). En los años siguientes, la investigación ha ido en ascenso.

Hunston (2002) resumió que los corpus se podían usar en exámenes para los siguientes propósitos:

- como un repositorio de exámenes,
- desarrollar materiales de pruebas,
- optimizar procedimientos de pruebas,
- mejorar la calidad de la corrección de pruebas,
- validar pruebas,
- estandarizar pruebas.

Dadas sus características, los corpus pueden tener un papel importante en los exámenes. Cabe destacar una serie de virtudes. En primer lugar, los corpus guardan gran cantidad de datos, lo cual posibilita no repetirse en la propuesta de textos. En segundo lugar, los datos de corpus son reales, vivos y convincentes, con lo que responden a aquella necesidad que subyace siempre, de una manera u otra, en aquél que se examina, a saber, ser competente en una situación comunicativa real. No hace falta que los profesores creen ejemplos. Los ejercicios diseñados con un corpus tienen autenticidad. Los profesores se encargan de analizar el nivel de dificultad y seleccionar muestras convenientes según su experiencia de enseñanza y su conocimiento sobre el corpus. En tercer lugar, los corpus anotados y los corpus de aprendices, pueden ayudar a los evaluadores a conocer informaciones importantes, como aquellos aspectos lingüísticos que un aprendiz encuentre especialmente farragosos, su distribución en los exámenes anteriores, etc. Por último, el listado de

frecuencias puede ayudar a decidir qué léxico y estructuras van a aparecer en el examen. Además de todo lo dicho, cabe añadir que se está estudiando su uso en nuevos ámbitos como los sistemas de evaluación automáticos.

Sin embargo, los corpus tienen sus limitaciones. Cada corpus se compila por diferentes motivos. Al seleccionarlos, hay que evaluar bien si las muestras son representativas y convenientes.

### ***3.1.4 El uso de corpus de aprendices de lengua segunda en la enseñanza de lenguas***

Uno de los desarrollos recientes es el uso de corpus de aprendices de lengua segunda. El primero fue el *Longman Learners' corpus*. Poco a poco se fueron construyendo más corpus de este tipo. La aparición del corpus *International Corpus of Learner English* (ICLE) a principios de los 90 simbolizaba el comienzo de los grandes corpus de aprendices (Granger et al., 2002).

En la creación de este tipo de corpus, primero hay que considerar las características de los alumnos que abarcan parámetros como la lengua materna, edad, dominio de otras lenguas extranjeras, ambiente de estudio, etc. Además, como los corpus pueden ser guardados a largo plazo, si las condiciones lo permiten, se puede seguir el estudio de los aprendices para poder describir y analizar todo el proceso de desarrollo de la interlengua.

Con este tipo de corpus, se suelen hacer comparaciones lingüísticas entre los alumnos de L2 y los hablantes nativos, entre los aprendices de diferentes países, entre los alumnos de distintos niveles, entre el idioma popular y el idioma para algún uso especializado, entre la lengua escrita y la lengua hablada de los aprendices, etc. Granger (2002) estaba convencida de que este tipo de comparación y análisis podía suministrar recursos e informaciones muy valiosas para la enseñanza de lenguas.

Entre sus exitosas aplicaciones se cuenta la posibilidad de ayudar a detectar las peculiaridades debidas a las interferencias de la lengua materna, sintetizar tipos de errores comunes cometidos por los alumnos de distintos niveles, así como a dirigir la práctica de la enseñanza y mejorar su calidad.

### ***3.2 La experiencia del laboratorio LLI-UAM***

#### ***3.2.1 Las ventajas y desventajas de corpus de habla espontánea en la enseñanza de lenguas***

En la Introducción, se ha indicado que el laboratorio LLI-UAM ha compilado varios corpus orales para diferentes aplicaciones. Entre ellos, hay corpus de diferentes idiomas, corpus de hablantes nativos y no nativos, corpus de lenguaje infantil, etc. Este laboratorio lleva años investigando, practicando y promoviendo la adaptación de corpus de habla espontánea a la enseñanza de lenguas extranjeras.

Conforme a su experiencia, ha logrado sintetizar y poner en claro las ventajas y desventajas del uso de corpus en la enseñanza. Ningún corpus puede ser completamente representativo de una lengua y contener todas las situaciones comunicativas o todas estructuras de la lengua. Debido a la diferencia del tamaño de unos corpus y otros, a veces las informaciones suministradas por algunos corpus están desordenadas, por lo que es imprescindible un filtro. Además, pueden incluir errores de expresión y gramática, palabras ofensivas, o convenciones de transcripción difíciles de entender, etc. Sobre todo, cuando se trata de un corpus de habla espontánea, las disfluencias, las incorrecciones propias de la oralidad como abreviaciones, cambios de acento particulares, elipsis de palabras... dificultan mucho su aplicación a la enseñanza.

Para evitar todo esto, los expertos del LLI-UAM ponen énfasis en un empleo pedagógico del corpus para su uso en la enseñanza. Es decir,

al presentar a los alumnos el uso real de la lengua, también hace falta advertirles acerca de las incorrecciones propias del habla descuidada. Siempre se recomienda que el profesor pre-edite el corpus manualmente y se familiarice con el corpus (principios de diseño, gramática, contenido comunicativo y léxico, etc.) y las herramientas, de manera que luego pueda enriquecer el uso del corpus con sus explicaciones (Campillos et al., 2010).

No obstante, las ventajas son múltiples. La principal es la autenticidad: los corpus ofrecen materiales auténticos procedentes del uso actual de la lengua y no del conocimiento intuitivo de un hablante peculiar o de la norma académica.

Además, la lengua aparece integrada en el contexto discursivo, por lo que, en el caso de la lengua oral, los alumnos pueden entender mejor algunas formas de expresión de la dicha lengua oral y su fraseología, sus peculiaridades estilísticas (vacilaciones, reformulaciones, elipsis), textuales (aperturas y cierres conversacionales) o pragmáticas (marcadores discursivos, marcas de cortesía), así como las características principales de la lengua hablada (Campillos et al. [en línea]). A los alumnos que por diferentes motivos no pueden ir al país nativo, les faltan medios para enfrentarse a situaciones comunicativas reales. En este caso, un corpus de habla espontánea contribuye a mostrarles las interacciones reales entre hablantes nativos en situaciones lingüísticas auténticas.

### ***3.2.2 Adaptación de C-ORAL-ROM a la enseñanza de ELE***

Dado su carácter oral, la adaptación de C-ORAL-ROM a la enseñanza de ELE tiene un objetivo concreto: su uso en el desarrollo de la comprensión auditiva.

La metodología que ha adoptado abarca los siguientes procedimientos.



1. Selección de muestras. Se ha tratado de que la información aportada por este corpus fuera adecuada para la didáctica de ELE. La metodología se concreta en el apartado 3.2.2.1.

2. Planteamiento de un marco teórico sobre una gramática pedagógica. Se inició con la elección de contenidos gramaticales, categorías léxicas y exponentes funcionales. La tabla 3.2, 3.3, y 3.4 nos presentan respectivamente el contenido.

1. Artículo 2. Sustantivo 3. Adjetivo 4. Determinantes y pronombres demostrativos 5. Determinantes y pronombres posesivos 6. Determinantes y pronombres indefinidos 7. Determinantes y pronombres numerales 8. Determinantes y pronombres interrogativos 9. Pronombres personales 10. Pronombres y adverbios relativos 11. Tiempos y modos verbales: 11.1 Indicativo: presente, futuro, pretérito perfecto simple, pluscuamperfecto 11.2 Imperativo 11.3 Condicional 11.4 Subjuntivo: presente, pret. imperfecto, pret. perfecto, pluscuamperfecto 11.5 Formas no personales: infinitivo, participio, gerundio	11.6 Perífrasis 11.7 La pasiva 11.8 Ser y estar 12. Construcciones con <i>se</i> 13. Preposiciones y locuciones prepositivas. Régimen preposicional 14. Adverbios y locuciones adverbiales 15. Referencias o marcadores espaciales y temporales 16. Coordinación. Nexos 17. Subordinación. Nexos y tipos de oraciones: 17.1 Sustantivas 17.2 Adjetivas o relativas 17.3 Adverbiales 18. Formación de palabras 19. Marcadores discursivos 20. Ortografía, acento gráfico y puntuación. 21. Siglas
---	---

Tabla 3.2 Contenidos gramaticales

<p><b>1. Nociones</b></p> <p>1.1 Ser y existir</p> <p>1.2 Cantidad</p> <p>1.3 Tiempo</p> <p>1.4 Espacio y lugar</p> <p>1.5 Relaciones entre eventos o sucesos</p> <p>1.5.1 Condición</p> <p>1.5.2 Concesión</p> <p>1.5.3 Causa</p> <p>1.5.4 Consecuencia</p> <p>1.5.5 Finalidad</p> <p>1.6 Modo</p> <p>1.7 Comparación</p> <p>1.8 Propiedad y posesión</p> <p>1.9 Intensidad y exclamación</p> <p>1.10 Impersonalidad</p> <p><b>2. Funciones comunicativas</b></p> <p>2.1 Usos sociales:</p> <p>2.1.1 Saludos y despedidas. Presentaciones</p> <p>2.1.2 Invitaciones. Concertar citas</p> <p>2.1.3 Dar las gracias</p> <p>2.1.4 Pedir disculpas</p> <p>2.1.5 Fórmulas sociales: brindis, comidas, condolencias, tratamiento por teléfono</p> <p>2.2 Estados físicos y emocionales:</p> <p>2.2.1 Gustos y preferencias</p> <p>2.2.2 Sentimientos y estados anímicos</p> <p>2.2.3 Deseos</p> <p>2.2.4 Estados físicos</p>	<p>2.3 Actitudes y conocimiento</p> <p>2.3.1 Posibilidad e imposibilidad. Capacidad e incapacidad</p> <p>2.3.2 Grado de certeza. Grado de probabilidad e hipótesis</p> <p>2.3.3 Conocimiento, recuerdo y olvido</p> <p>2.3.4 Opinión</p> <p>2.3.5 Acuerdo y desacuerdo</p> <p>2.3.6 Obligación y necesidad</p> <p>2.4 Influencia:</p> <p>2.4.1 Consejos, advertencias y recomendaciones</p> <p>2.4.2 Sugerencias y propuestas</p> <p>2.4.3 Peticiones</p> <p>2.4.4 Quejas, reclamaciones</p> <p>2.4.5 Regañar. Amenazar e insultar</p> <p>2.4.6 Animar a la acción</p> <p>2.4.7 Promesas, compromisos y juramentos</p> <p>2.4.8 Instrucciones, órdenes, prohibiciones.</p> <p>2.4.9 Permitir y pedir permiso.</p> <p>2.5 Comunicación:</p> <p>2.5.1 Destrezas de comunicación oral</p> <p>2.5.2 Organización discursiva</p> <p>2.5.3 Control del lenguaje</p> <p>2.5.4 Discurso referido</p>
---	--

Tabla 3.3 Nociones y funciones comunicativas

1. Identificaciones personales 2. La profesión y el trabajo 3. La religión y la ideología 4. La familia 5. Carácter, valores y estado de ánimo 6. La casa y el hogar 7. Tiempo libre y ocio 8. La naturaleza 9. El tiempo y el clima 10. El arte y la cultura 11. Medios de comunicación 12. Deportes y espectáculos 13. Los viajes	14. La residencia y el alojamiento 15. El transporte 16. El cuerpo y la salud 17. La educación 18. La gastronomía 19. Comercio, negocios e industria 20. Instituciones públicas y civiles 21. Pesos, medidas y cantidades 22. El estado y la sociedad 23. Descripción de objetos 24. Ordenadores, Internet y tecnología 25. Hábitos y valores culturales 26. Fraseología
---	--

Tabla 3.4 Contenidos léxicos

3. Desarrollo de una herramienta de búsqueda. Dicha herramienta consiste en un sistema de concordancias con sonido: el usuario escribe la palabra o palabras que desea consultar y el sistema le muestra los ejemplos contenidos en el corpus (si no saca ningún fragmento es que el corpus no tiene esa expresión).

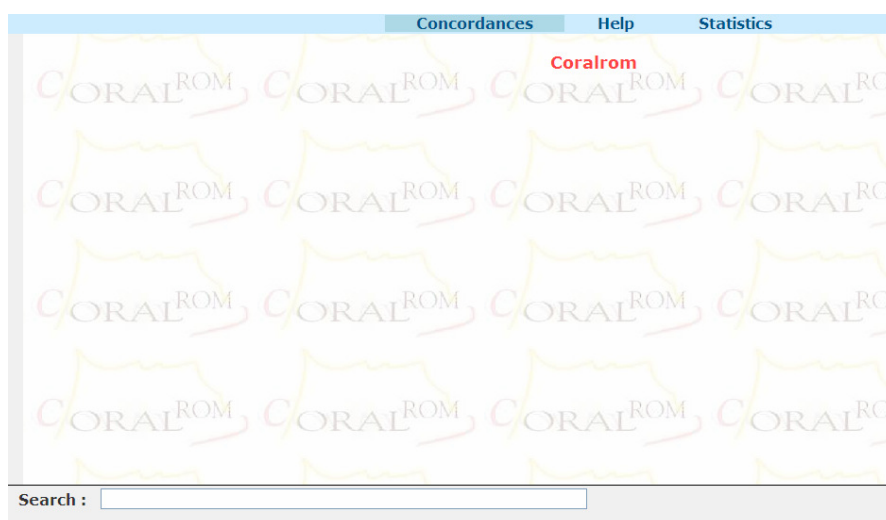


Figura 3.5 Herramienta de búsqueda de C-ORAL-ROM

En la figura 3.5, los profesores y alumnos pueden introducir el contenido que quieran buscar en el cajetín de *Search* y consultarlo. Esta herramienta de concordancias puede servir para ver y oír ejemplos del habla espontánea. Un sistema como este permite mucha libertad, pero tiene el inconveniente de que si el usuario no sabe muy bien lo que busca, los resultados pueden ser muy confusos. Como los aprendices de una lengua no dominan ni conocen muchos de los conceptos gramaticales o léxicos que quieren buscar, sería muy útil organizar las búsquedas por temas y representarlas de una manera ordenada. De esta manera, la herramienta serviría tanto para profesores (como complemento a sus explicaciones) como para alumnos (en tareas de autoaprendizaje).

### ***3.2.3 La experiencia de la elaboración de la serie Español Oral en Contexto***

Merece una especial mención el libro *Textos de Español Oral* de la serie *Español Oral en Contexto* (Campillos et al., 2010).

El material se acompaña de una herramienta de consulta en formato hipertextual para consultar y escuchar los contextos lingüísticos. Sirve como un complemento a los manuales existentes o a los

recursos propios del docente y proporciona, además, actividades diseñados para los usuarios. El principal objetivo para el que puede contribuir consiste en la mejora de la destreza de comprensión auditiva. Es conveniente principalmente para profesores o estudiantes de niveles intermedio-avanzado (B1-B2) o superior-perfeccionamiento (C1-C2). Como el nivel de dificultad es parecido a nuestro corpus C-ORAL-CHINA, sirve como modelo en nuestro futuro trabajo de explotación del corpus para la enseñanza de chino.

Este recurso didáctico incluye:

- Ejemplos de uso de habla real.
- Gran variedad de temas y situaciones comunicativas:
- Diversos géneros y tipos de discurso oral (monólogos, diálogos o conversaciones).
- Muestras del registro formal y del coloquial.

### ***3.2.3.1 Metodología y criterios de selección de documentos***

La selección de los textos ha sido manual. Todos los ejemplos se han seleccionado de los documentos de subcorpus de español de C-ORAL-ROM. En total, se ofrecen 200 fragmentos de entre 20 segundos y 3 minutos de duración, que suman un total de 6 horas de grabaciones para la práctica de la comprensión auditiva. Son más de 6.000 ejemplos presentes en más de 3.500 sonidos. Según ellos, la experiencia en la elaboración del recurso permite afirmar que dichos ejemplos recogen los elementos más fundamentales para el correcto aprendizaje de una lengua o, cuando menos, una amplia mayoría, si es que se tiene en cuenta la consideración de G. Aston sobre la preferencia de un corpus de entre 20.000 y 200.000 palabras para un contexto de enseñanza (tomado de Campillos et al., 2010: IX).

Básicamente, el proceso de la elaboración abarca las siguientes etapas:

- Escuchar el documento, valorar la adecuación de sus contenidos para la didáctica, y seleccionarlo. Los documentos que adolecen de mala calidad acústica, contenidos inconvenientes para la didáctica o temas no interesantes no han podido ser seleccionados.
- Leer detenidamente las transcripciones de los documentos (adaptadas a las convenciones de la lengua escrita) y describir sus contenidos gramaticales, léxicos, su registro o su velocidad de habla, etc.
- Clasificar el documento en un nivel de dificultad según el Marco Común Europeo de Referencia para las Lenguas.

Durante todo el proceso, hay dos puntos muy importantes. Uno es la selección de conversaciones dentro del ámbito informal. Dicho ámbito presenta un alto grado de elipsis de estructuras en el habla (del tema tratado, o de partes de la oración fundamentales como verbos o sujetos, que quedan implícitos en el contexto situacional). Junto a ello, en las conversaciones coloquiales existe una frecuente falta de planificación del discurso, que no solamente se refleja en una organización de las estructuras lingüísticas diferente a la del discurso escrito, sino también en la ausencia de un tema central que aglutine los motivos de la conversación. En efecto, muchas conversaciones coloquiales se han llevado a cabo para socializarse o mantener un vínculo con el interlocutor, más allá del interés de transmitir datos. Todos estos factores hay que tenerlos en cuenta en la búsqueda de ejemplos.

El otro punto relevante centra su atención sobre las situaciones comunicativas concretas del uso de la lengua. Las personas que utilizan un corpus no son los destinatarios originales de los textos que lo componen, y, precisamente por eso, la adaptación de los mismos a la enseñanza requiere un proceso de mediación pedagógica que procure reconstruir el contexto discursivo auténtico. La recreación

del contexto original por parte del estudiante será tanto más accesible cuanto mayor familiaridad tenga éste con el tema del texto y cuanto más interesante le resulte. Bajo este presupuesto, los autores han intentado enmarcar algunos textos extraídos en un tipo de actividad que pueda situar al alumno en una situación comunicativa semejante a la original (Campillos et al., 2010).

### ***3.2.3.2 Tipología de actividades***

Se han diseñado una serie de actividades. Principalmente son las siguientes:

1. Selección de información:

- Preguntas de verdadero / falso
- Rellenar huecos
- Rellenar tablas
- Completar información
- Marcar las casillas
- Ordenar los diálogos
- Ordenar elementos.

2. Transferencia de información:

- Con ilustraciones, se pueden realizar actividades como escuchar y seguir una ruta, marcar en un mapa qué ciudades se mencionan o los fenómenos climáticos de cada región.
- Dibujar un plano de una casa a partir de los datos escuchados.

3. Relación o distinción de información:

- Relacionar datos
- Distinguir la información escuchada y la que se ofrece en forma de texto:
- Corregir un texto

- De entre varias ideas clave que se presentan sobre un documento, marcar las que se escuchan o en qué orden aparecen

4. Sintetizar información: se trata de actividades de resumir los contenidos en un breve texto, empleando un título o completando una frase.

5. Reacción y respuesta:

- responder a preguntas de respuesta abierta o cerrada, de selección múltiple, etc.
- escuchar la grabación y formular un consejo.

6. Actividades de inferencia: en estos ejercicios, el hablante escucha una grabación y deduce los datos necesarios para realizar una tarea, por ejemplo: interpretar información o inferir datos desconocidos de la grabación.



## 4. El chino *putonghua*

### 4.1 *Visión global*

La República Popular de China está compuesta por cincuenta y seis grupos étnicos. Entre ellos, la etnia *Han*, que ocupa más del 90% de la población, es la más numerosa. El idioma hablado por los *Han* se llama *hanyu*. Es el idioma principal de los chinos. Según Ethnologue<sup>5</sup>, además de en China, también se habla en otros países como Singapur, Malasia, Indonesia, Mongolia, Tailandia, etc. Dependiendo del lugar donde se hable, también se denomina como *zhongwen*, *guoyu*, *huayu*, etc.

Hay otras lenguas en China con su propio sistema de escritura, como el mongol, el tibetano o el manchú. Principalmente, las hablan los grupos étnicos minoritarios.

La forma estandarizada del chino es *putonghua*. Es la lengua representativa de la China moderna. Pertenecce a la familia sino-tibetana. Es uno de los idiomas con más hablantes del mundo y una de las lenguas oficiales de las Naciones Unidas.

Jerry Norman, catedrático estadounidense de la lengua china de la Universidad de Washington, ha descrito el idioma chino en su libro *Chinese*:

Chinese is only one of a very few contemporary languages whose history is documented in an unbroken tradition extending back to the second millennium BC. At the same time, in its numerous dialectal forms, it has more speakers than any language spoken in the modern world (Norman, 2008: 27).

---

<sup>5</sup> “Chinese, mandarin”. En: *Ethnologue Languages of the World* [en línea]: <  
[http://www.ethnologue.com/show\\_language.asp?code=cmn](http://www.ethnologue.com/show_language.asp?code=cmn)> [consulta: 15 de febrero de 2011]

Es un idioma que tiene una larga historia y muchos dialectos. Si cada individuo habla su dialecto, es difícil comunicarse. Como lengua común, el *putonghua* facilita mucho el intercambio entre gente que habla diversos dialectos de diferentes zonas. El estudio del *putonghua* es obligatorio en los colegios y es el idioma de la administración pública y los medios de comunicación. Su posición está firmemente establecida.

En esta tesis, cuando se habla de la lengua china, se refiere a *putonghua*.

## ***4.2 Evolución del idioma chino***

La evolución de la lengua china y la historia del pueblo chino están muy relacionadas. Según las características de la transformación del chino, éste se puede dividir en distintas etapas. Pero, como el desarrollo de una lengua es gradual, es imposible indicar qué año exacto es el comienzo de una etapa histórica del chino. Los sinólogos chinos y extranjeros han planteado puntos de vista sobre la clasificación.

En esta tesis, vamos a seguir la conocida clasificación de Wang Li (1980) en su libro *Hanyu Shigao*.

**Chino arcaico** (antes del siglo III d.C. Entre siglo III y siglo IV, se da la etapa de transición.)

La escritura china en caparazones de tortugas y huesos de bóvidos, datados de la dinastía Shang, son los primeros caracteres maduros.

Durante el período de Primavera y Otoño y el de los Reinos Combatientes (770-221 a. C.), el chino cambia mucho. La anexión y unión de las tribus o reinos hacía que muchos dialectos pequeños locales se mezclaran y se formaran dialectos grandes. La literatura llegó a una etapa cumbre. Apareció el fenómeno de debate de las

Cien Escuelas del pensamiento (Gao Jianping et al., 1980). Aunque las obras literarias de los grandes pensadores tenían características locales, poco a poco se desarrollaba una lengua literaria unificada denominada *yayan* (lengua elegante).

Qin Shihuang, el primer emperador chino, unificó toda China en el año 221 a.C. China pasó a ser un estado centralista. Se aplicó una medida llamada *shutongwen* para unificar la escritura de todo el país. Asimismo, se estableció que se usara *xiaozhuan* como estándar de los caracteres.

El aspecto más positivo de *shutongwen* consistió en que se garantizó el contacto sostenible y la transmisión de información entre el gobierno central y las distintas provincias, en todos aspectos. Un carácter podía tener múltiples pronunciaciones, pero su forma de escritura era estable y unificada. Tal desarrollo supone la base para la unificación de la lengua. En las dinastías siguientes, se mantenía dicha tradición que favorece notablemente la herencia histórica de la cultura. Por esta razón, los chinos de hoy en día pueden entender las obras literarias clásicas.

Según Wang Li (1980), en esta etapa las características generales son: en las oraciones enunciativas, no se usa la cópula; en las oraciones interrogativas, el objeto pronominal se coloca delante del verbo; hay dos tonos entrantes (uno de los cuatro tonos de la pronunciación del chino clásico, aún existente en ciertos dialectos), etc.

**Chino antiguo** (siglo IV - siglo XII d.C. Entre siglo XII y siglo 13, se da la etapa de transición.)

En comparación con la etapa anterior, la estructura gramatical se desarrolló mucho. Hay algunos fenómenos lingüísticos muy representativos en esta época. Por ejemplo, el nacimiento de “是 (el verbo copulativo)”, la oración pasiva con “被”, el nacimiento de la

partícula “了” y “着”, el nacimiento de caracteres con tono descendente, etc (Gao Jianping et al., 1980). Además, con la estabilidad del significado y la función de las palabras, poco a poco aparecía la diferenciación de categorías de palabras.

**Chino moderno** (siglo 13- siglo 19. Entre 1840 y 1919, se da la etapa de transición. )

En la historia, la centralización política y el separatismo durante largo plazo mantenían la unificación del chino, sobre todo la de la escritura. Como la población de la etnia *Han* crecía mucho y la superficie de distribución se extendía, en esta época, se formaban los siete grandes grupos de dialectos: dialectos del norte, Wu, Xiang, Gan, Kejia o Hakka, Min y Yue. Con el cambio de las condiciones políticas, económicas y culturales, los dialectos del norte cobran cada vez más importancia.

En la dinastía Ming, se comenzó a hablar *guanhua* (lengua oficial). Tomaba el dialecto de Nanjing como referencia. Después del traslado de la capital de Nanjing a Beijing, se mezclaban el dialecto de Nanjing y el de Beijing. En 1728, el emperador Yong Zheng de la dinastía Qing publicó un decreto imperial, creando las academias de *zhengyin* (correcta pronunciación) para promover el *guanhua* de Beijing en toda China. En 1909, se estableció el *guoyu* (lengua nacional) en sustitución del *guanhua* (Wen Songhui, [en línea]).

En esta época, surgieron muchos escritores famosos. Utilizando la lengua de manera variada y viva, escribieron numerosas obras literarias. Su estructura del lenguaje y el empleo de ricas formas de expresión sientan las bases el desarrollo del chino contemporáneo.

**Chino contemporáneo** (1919-actualidad)

Durante las dinastías Tang y Song surgió el *baihuawen* (una forma de chino más cercana a la lengua hablada). Y también aparecieron novelas escritas en *baihuawen*. Sin embargo, la posición dominante

de *wenyanwen* (chino clásico) no fue modificada ni desafiada. El *wenyanwen* se fue desconectando del lenguaje hablado poco a poco, con lo cual ya no se adaptaba al desarrollo de la sociedad y la lengua popular.

Alrededor del Movimiento del 4 de Mayo de 1919, se inició el movimiento a favor del uso generalizado del *baihuawen* a través de la revista “Nueva Juventud” (*Xinqinnian*), provocando un debate entre los defensores del *baihuawen* y los del *wenyanwen*. Este debate formó parte del Movimiento de la Nueva Cultura. No sólo fue una reforma dentro de la lengua, sino que también estaba estrechamente relacionada con la reforma de la conciencia de pensamiento y la modernización del país. La confrontación entre el *baihuawen* y el *wenyanwen* reflejó en gran medida el choque entre distintos sistemas de valores e ideologías sociales. Finalmente la época del *baihuawen* finalizó, justo en el momento histórico donde se produjo la transición de China desde la sociedad feudal a la democrática.

A partir de entonces, el chino empezaba a absorber el léxico y la gramática del occidente.

En 1932, el Ministerio de Educación definió el dialecto de Beijing como la pronunciación de la lengua nacional (Wen Songhui, [en línea]).

En 1955, se celebró la Conferencia Académica de la Normalización del Chino Moderno en la que se planteaba oficialmente la definición de *putonghua*. Toma la fonética de Beijing como estándar, el habla del norte como dialecto básico y obras modernas en *baihuawen* (el escrito en lengua moderna) como base para las reglas gramaticales (VV.AA., 2006).

En febrero de 1958, en la Asamblea Popular Nacional de China, se aprobó el esquema del alfabeto fonético del chino, denominado como *hanyu pinyin*. En el apartado 4.4, los comentaremos detalladamente.

### 4.3 Caracteres chinos

Uno de los aspectos más llamativos y peculiares de la lengua china es su escritura. Se trata de una de las escrituras más antiguas del mundo. Fue adoptada por otras lenguas asiáticas, como por ejemplo, el japonés, el vietnamita y el coreano. En chino, los caracteres se llaman *hanzi* (汉字). Son combinaciones de trazos evolucionados sobre la base de los primeros pictogramas.

#### 4.3.1 El número de caracteres chinos

A lo largo de la historia, la escritura del chino ha experimentado muchos cambios y reformas. Como resultado de esta evolución y acumulación, si se tienen en cuenta todo tipo de formas variantes, cada vez hay más caracteres.

La tabla 4.1 nos muestra el número de caracteres chinos en los diccionarios más representativos en cada época histórica.

Fecha	Dinastía	Diccionario	Número de caracteres
100	Han de Este	<i>Shuōwén jiězì</i>	9,353
543	Liang	<i>Yùpiān</i>	12,158
1011	Song	<i>Guǎngyùn</i>	26,194
1039	Song	<i>Jíyùn</i>	53,525
1615	Ming	<i>Zìhuì</i>	33,179
1716	Qing	<i>Kāngxī zìdiǎn</i>	47,035
1916	Minguo	<i>Zhōnghuá dà zìdiǎn</i>	48,000
1994	República Popular de China	<i>Zhōnghuá zìhǎi</i>	85,568

Tabla 4.1 Número de caracteres en diccionarios chinos (Norman, 2008)

Esta cifra es muy sorprendente. En *Zhōnghuá zìhǎi*, editado hace 17 años, se recogen más de 85,000 caracteres. Aquí cabe señalar que de todos estos, unos 3,500 (VV.AA., [en línea]) son los habituales. Con dominarlos, una persona ya puede leer la prensa.

### **4.3.2 Origen y evolución**

Los caracteres chinos cuentan con una historia muy larga.

En las Restos Arqueólogos de la Edad Neolítica, de unos 6,000 de antigüedad, se han descubierto signos simples y regulares inscriptos en cerámicas, los cuales han sido considerados por arqueólogos y filólogos como la forma embrionaria.

Hace más de 3,000 años, las inscripciones en caparazones de tortuga y huesos de animal, procedentes de la dinastía Shang, han sido considerados como escritura madura y estructurada.

Gráficamente, la escritura china ha experimentado una evolución gradual de pictogramas a logogramas cuadrangulares compuestos por trazos, de formas complejas a simples. Los estilos de caligrafía más conocidos son *jiaguwen* (inscripciones en caparazones de tortuga y huesos de animal), *jin* (inscripciones en antiguas vasijas de bronce), *zhuanshu* (escritura de sellos), *lishu* (escritura oficial), *caoshu* (escritura cursiva), *kaishu* (escritura regular) y *xingshu* (escritura corrida), etc. Entre ellos, *kaishu* nació a finales de la dinastía Han. Presenta una forma cuadrada, la cual se modela y se utiliza hasta la actualidad.

Después de la fundación de la República Popular de China, el gobierno chino empezó un proceso de reforma con el objetivo de simplificar los caracteres, sustituyendo la escritura compleja por una forma de escritura más sencilla y compuesta por menos trazos. Así que las grafías son más simples y abstractas. Resultarán más fáciles de recordar y se empleará menos tiempo a la hora de escribirlos. Pero,

en algunas zonas como Taiwan, Hong Kong y Macao, aun utilizan la escritura sin simplificarse, conocida como la tradicional.

### 4.3.3 Estructura

En el diccionario *Shuōwén jiězì*, redactado por Xu Shen en el siglo I de nuestra era, se introducen seis clases etimológicas (en chino, *liushu*) sobre el origen de cada carácter. Hoy en día, el estudio sobre la estructura de los caracteres chinos no mantiene estas seis clases. Generalmente, se cree que entre las seis, cuatro explican la formación de caracteres.

(1) Pictograma (*xiangxing*). Describe según las características físicas del objeto. Por ejemplo, el carácter “日 (sol)” imita la forma del sol.

(2) Caracteres indicadores (*zhishi*). Es la forma para demostrar las cosas abstractas. Por ejemplo, “上 (encima, arriba)”、“下 (debajo, abajo)” son respectivamente trazos añadidos encima y debajo de una línea transversal, indicando la posición relativa de los dos caracteres.

3) Ideograma (*huiyi*). Une dos o más signos para formar un nuevo carácter. Por ejemplo, “日 (sol)” y “月(luna)” forman “明(brillante, claro)”.

(4) Fonograma (*xingsheng*). Está compuesto por dos partes: un radical que aporta el significado semántico y una parte fonética que indica la pronunciación. Por ejemplo, para el carácter “櫻(cerezo, yīng)”, “木” significa que es un tipo de árbol y “嬰(yīng)” determina su pronunciación.

### 4.3.4 Las reglas del orden de los trazos de la escritura

En cuanto a la manera de componer y leer un texto escrito, los



caracteres chinos se escriben de izquierda a derecha, de arriba hacia abajo como la escritura occidental. Sin embargo, a diferencia del español, los caracteres se colocan uno tras otro sin espacios entre medio, incluso cuando se trate de palabras distintas.

A continuación se enumeran unas reglas básicas del orden de los trazos en la escritura china (Li Xingjian, 2004).

Reglas fundamentales:

1. Primero el trazo central y luego los laterales. 水
2. Primero los trazos horizontales y luego los verticales. 十
3. Primero los trazos superiores y luego los inferiores. 前
4. Primero los trazos del lado izquierdo y luego los del derecho. 作
5. Primero los trazos exteriores y luego los interiores, si hay trazos que rodean. 国
6. Primero los trazos exteriores, luego los interiores, y al final el cierre si es como una caja cerrada. 国
7. Primero el trazo derecho y luego el izquierdo cuando dos trazos curvos se cruzan en forma de "x". 风

Reglas adicionales:

1. Si el punto está en la parte superior o la parte superior del lado izquierdo, primero el punto. 主
2. Si el punto está en la parte superior del lado derecho o dentro del carácter, el punto se escribe el último. 瓦, 我
3. En caracteres con elementos que rodean otros elementos, el orden depende de:
  - Si hay trazos que rodean la parte superior y la parte derecha o la parte superior y la parte izquierda, primero los trazos exteriores y luego los interiores. 厅, 庆
  - Si hay trazos que rodean la parte izquierda y la parte inferior, primero los trazos interiores y luego los exteriores. 远, 建, 廷
  - En palabras con un elemento como un cuadrado, pero sin el

lado superior, primero los trazos interiores y luego los exteriores. ㄨ, ㄩ

- En palabras con un elemento como un cuadrado, pero sin el lado inferior, primero los trazos exteriores y luego los interiores. 同, 用, 风
- En palabras con un elemento como un cuadrado, pero sin el lado derecho, primero los trazos superiores, luego los interiores, y al final la parte inferior de la izquierda. 巨, 匠, 区

#### 4.4 Hanyu pinyin

Para la transcripción fonética de los caracteres chinos se usa una romanización de los sonidos: el *pinyin*. Está aceptado por el gobierno chino y por la comunidad internacional. El *pinyin* tiene sus propias normas de pronunciación, las letras no se pronuncian como en castellano ni en inglés.

##### 4.4.1 Evolución

En China, antes no había *pinyin*. Para marcar el signo fonético, se utilizaban los métodos *zhiyin* o *fanqie*. *Zhiyin*, es el método tradicional de indicar la pronunciación de un carácter chino citando otro de la misma pronunciación. Pero, si el otro también es un carácter complicado o raramente usado, resulta evidentemente complicado de comprender. El método *fanqie*, consistía en indicar la pronunciación por medio de otros dos caracteres, de los cuales el primero proporciona la consonante y el segundo la vocal (con o sin nasal final) y el tono (Sun Yizhen, 1999). Por ejemplo, el carácter “胜” (la pronunciación es *shèng*), se señala como “师”(shī) y “正”(zhèng), lo que significa una combinación de la consonante *sh* de *shī* y la vocal más nasal *eng* y el cuarto tono de *zhèng*. Este método tiene una desventaja. Para manejarlo, hay que conocer primero más o menos 1000 caracteres (Hu Jicheng, [en línea]), lo cual también

supone una carga.

Bajo estas circunstancias, se necesitaba urgentemente una serie de signos fáciles y prácticos, con el fin de compensar los defectos de los métodos tradicionales.

Al final de la dinastía Ming, los misioneros occidentales llegaron a China. En 1605, en *Xizi Qiji (Wonder of Western Writing)* (Yin Binyong, [en línea]), publicado en Beijing por Matteo Ricci (jesuita italiano 1552-1610), se introdujo el alfabeto fonético latino para intentar una transcripción fonética del chino, con el objetivo de facilitar a los extranjeros el aprendizaje del idioma y la escritura. Este intento inspiraba a muchos expertos de generaciones posteriores a intentar diseñar y promover un alfabeto romanizado para la lengua china.

Antes del establecimiento del esquema oficial del alfabeto fonético del chino, había tres normas diseñadas por los chinos que tenían influencia. Se trataba de: *zhuyin zimu* (publicado en 1918), *guoyu luomazi* (publicado en 1926) y *ladinghua xinwenzi* (publicado en 1931) (Hu Jicheng, [en línea]). Debido a las condiciones históricas, tenían sus limitaciones. Después de la fundación de la República Popular de China, con el objetivo de promover el *putonghua*, la lengua común, y absorbiendo las ventajas de los esquemas anteriores fruto del esfuerzo de los antepasados, se elaboró el esquema del alfabeto fonético del chino que fue aprobado en febrero de 1958 en la Asamblea Popular Nacional de China, denominado como *hanyu pinyin*.

#### **4.4.2 Características**

Está compuesto por la tabla de alfabeto fonético chino, tabla de consonante inicial, tabla de vocal, signos de tono y signos divisorios silábicos. Tras considerar y rechazar varias propuestas, finalmente, en este esquema, se escoge el alfabeto latino. Es un tipo

de letras internacional; es fácil de escribir y leer; y permite analizar y describir la pronunciación de manera minuciosa. En total, hay 26 letras, entre ellas, 21 son consonantes. Su orden alfabético también se conforma con el orden tradicional de las letras latinas, desde *a* hasta *z*.

Como 26 letras no son suficientes, se aplican unos métodos, por ejemplo, la fórmula de una nueva a través de dos letras individuales, como *zh*, *ch*, *sh* y *ng*, o el uso de signos adicionales, como el caso de *ü*. Además, el chino tiene otra peculiaridad: es un idioma tonal. El tono juega un papel muy importante en distinguir la semántica de morfemas cuasihomónimos. La pronunciación incorrecta puede provocar malentendidos y confusiones. Existen cuatro tonos principales (Zhao Shiyu, 1999):

- El primer tono: alto y sostenido, representado por signo tonal “\_”
- El segundo tono: ascendentes de medio a alto, representado por signo tonal “/”
- El tercer tono: descendente y ascendente, representado por signo tonal “v”
- El cuarto tono: descendente de alto a bajo, representado por signo tonal “\”
- Finalmente, el tono neutro (también llamado tono ligero): corto y ligero, un fenómeno fónico de modificación de tono, sin signo tonal

#### **4.4.3 Uso**

Después de su establecimiento, el uso del *pinyin* se populariza y se extiende ampliamente. Ahora los manuales, la prensa, los diccionarios, los envases y embalajes, carreteras, etc, tienen marcado el *pinyin*.

Este alfabeto fonético es el que se usa en la etapa inicial del *putonghua* para conocer la pronunciación de cada nuevo carácter que aprenden. En la enseñanza de chino al extranjero, constituye una herramienta auxiliar para el aprendizaje.

Gracias a las nuevas tecnologías informáticas, los chinos pueden usar el *pinyin* para escribir con mucha facilidad en ordenadores o en teléfonos móviles. Al escribir, no es necesario aislar el radical de cada carácter, como la manera clásica de buscar una palabra en un diccionario de papel chino. Es mucho más sencillo: Se escribe en *pinyin* en el teclado y en la pantalla aparecen en caracteres las distintas palabras asociadas a esa pronunciación. Al encontrar la opción deseada, el usuario la selecciona y automáticamente aparece en pantalla el carácter correspondiente. Como esta tecnología cada vez más inteligente, en algunos sistemas avanzados, para escribir las palabras más habituales, con sólo escribir la consonante inicial de cada carácter, ya aparecen las opciones, lo cual ha aumentado considerablemente la rapidez con la que un usuario puede escribir un texto en chino.

Asimismo, ha servido como base para la reforma de la escritura de algunas étnias minoritarias.

Finalmente, se utiliza en áreas donde no es fácil de usar caracteres chinos, por ejemplo, códigos telegráficos, braille para ciegos, lengua de signos para sordos, etc.

#### **4.5 Dialectos**

Dentro de la etnia *Han*, los chinos suelen llamar dialectos (方言, *fangyan*) al habla local de diferentes zonas del país. Estos dialectos regionales conviven en la República Popular de China con otras lenguas como el tibetano o el uigur. Estas lenguas se hablan por otras etnias minoritarias, y tienen su propio sistema de escritura.

Los dialectos pueden clasificarse en siete grandes grupos (Yuan Jiahua, 2001): dialectos del norte (*beifang fangyan*, 北方方言), dialectos de Wu (*Wu fangyan*, 吴方言), dialectos de Xiang (*Xiang fangyan*, 湘方言), dialectos de Gan (*Gan fangyan*, 赣方言), dialectos Kejia o Hakka (*Kejia Fangjian*, 客家方言), dialectos de Min (*Min fangyan*, 闽方言) y dialectos de Yue (*Yue fangyan*, 粤方言). Entre ellos, los primeros cubren una extensión más amplia y tienen mayor cantidad de hablantes. Alrededor del 70% de la población *Han* los habla. Además, como hemos comentado en el apartado 4.1, son la base de *putonghua*.

Los dialectos chinos son muy complicados. Tienen sus propios fondos históricos, sistema de pronunciación y características de léxico y gramática. Entre diferentes grupos, son mutuamente ininteligibles. Incluso dentro del mismo grupo, los hablantes pueden no entenderse unos a otros. Pero, todos usan los mismos caracteres chinos como escritura. “China has always had a uniform written language which is logographic. People who cannot understand each others’ speech can still read the same written language provided that they are educated (Comrie, 1990: 813).” Aunque puede que fonéticamente no se entienda, la comprensión semántica es igual porque la gente puede leer los mismos caracteres.

Como el corpus C-ORAL-CHINA está destinado a la enseñanza del idioma chino, hemos pedido a todos los participantes que en la grabación hablaran el *putonghua*, aunque unos de ellos también pueden hablar dialectos de su pueblo. Con respecto a esto, vamos a hablar con más detalles en el Capítulo 5.

#### ***4.6 Lengua oral y escrita***

Tanto la oralidad como el surgimiento de la escritura a partir de la oralidad son necesarias para la evolución de la conciencia. (Walter Ong, 1987)

La lengua oral está formada por sonidos, y la lengua escrita, por caracteres. Cuando hablamos, podemos aprovechar el cambio de la velocidad, la entonación, el lenguaje corporal y el entorno para expresar la idea. La lengua oral es más natural y espontánea, con oraciones más sencillas. Es inevitable que haya repeticiones, pausas, apoyos vocálicos, etc. La lengua escrita es más estructurada y más exigente con el uso de palabras. Trata de excluir palabras ociosas.

Como comentamos en el apartado 4.2, en la historia, debido a que el *wenyanwen* se desvinculaba de la lengua oral, se inició el movimiento a fin de preconizar el *baihuawen*, más popular y fácil de entender. En el chino contemporáneo, esencialmente, la lengua oral y la lengua escrita son coherentes entre sí, pero también presentan notables diferencias.

En un contexto, el intercambio es inmediato, por eso, la lengua oral es una acción inmediata y se puede corregir en seguida. En cambio, como la lengua escrita puede sobrepasar el límite de tiempo y espacio, es imprescindible la reflexión y el refinamiento. Puesto que son dos formas distintas de comunicación, el vocabulario que se usa es distinto. En un país tan grande como China que tiene muchos dialectos, tanto en la historia como en la actualidad, hay grandes diferencias entre el vocabulario oral y el escrito.

Por ejemplo, el carácter “信” y el “函”. El primero tiene mucho matiz oral. Es más corriente y sencillo. El segundo tiene matiz escrito. Es más culto y formal.

<b>Número de sílabas</b>			
	<b>Monosílabos</b>	<b>Bisílabos</b>	<b>Polisílabos</b>
Cantidad de palabras de la lengua oral	102	545	155
Porcentaje	12.72%	67.95%	19.33%
Cantidad de palabras de la lengua escrita	1604	2187	30
Porcentaje	41.98%	57.24%	0.78%

Tabla 4.2 Contabilidad de número de sílabas de la lengua oral y la escrita

<b>Número de morfemas</b>		
	<b>Palabra de un solo morfema</b>	<b>Palabra compuesta</b>
Cantidad de palabras de la lengua oral	113	689
Porcentaje	14.09%	85.91%
Cantidad de palabras de la lengua escrita	1774	2047
Porcentaje	46.43%	53.57%

Tabla 4.3 Contabilidad de número de morfemas de la lengua oral y la escrita



Número de significados		
	Palabras que sólo tienen un sentido	Polisémicas
Cantidad de palabras de la lengua oral	706	96
porcentaje	88.03%	11.97%
Cantidad de palabras de la lengua escrita	3237	584
Porcentaje	84.72%	15.28%

Tabla 4.4 Contabilidad de número de significados de la lengua oral y la escrita (Cao Wei, 2003)

En la tabla 4.2, 4.3 y 4.4, Cao Wei ha hecho una contabilidad de cantidad entre palabras de la lengua oral (se indica con 〈口〉) y palabras de la escrita (se indica con 〈书〉), palabras de *wenyan*) en el Diccionario de Chino Contemporáneo (versión 1983).

Los datos nos demuestran que en el chino contemporáneo, como las palabras orales se forman a base de *baihuawen*, hay más palabras compuestas que palabras de un solo morfema. Hay más bisílabos y polisílabos que monosílabos. Como no tiene una historia larga, ni la divulgación por escrito, hay más palabras que sólo tienen un sentido que polisémicas.

	Dialectos	Lengua oral	Lengua escrita
子	201	122	15
儿	203	145	1
头	75	29	1

Tabla 4.5 Contabilidad de palabras que llevan tres sufijos habituales en dialectos, lengua oral y lengua escrita

La tabla 4.5 es una comparación de las palabras que llevan tres

sufijos habituales de nombres 子, 儿 y 头 respectivamente en dialectos (se indica con 〈方〉), lengua oral (se indica con 〈口〉) y lengua escrita (se indica con 〈书〉) en el mismo diccionario. Se nota que obviamente estos sufijos se usan mucho en dialectos y lengua oral, y muy poco en la lengua escrita.

En torno a la gramática, la lengua oral no se atiende tanto al convencionalismo, hay más oraciones cortas y elípticas. A veces no concuerdan estrictamente con los principios gramaticales.

Además, la lengua oral se basa en el sonido, así que se usan más interjecciones, palabras de forma inacentuada que cumplen funciones gramaticales de estructura, de tiempo, o de modo, la duplicación de palabras, oraciones exclamativas, oraciones interrogativas, etc. Y la lengua escrita se concentra más en el sentido del carácter. Algunos dichos clásicos de *wenyanwen* pueden aparecer en ella.

Sin embargo, la lengua oral y la escrita no son contradictorias. También se influyen mutuamente. Los componentes de la lengua escrita también entran en la lengua oral. Los componentes de la lengua oral también se absorben en la lengua escrita para que esté más viva y rica. En la enseñanza del chino como segunda lengua, se puede motivar a los alumnos a conocer las relaciones mutuas entre ellas y peculiaridades de cada una. Por medio de clases de lectura y redacción y clases de lengua oral, se intenta elevar de manera eficaz la capacidad de comunicación escrita y oral de los alumnos.

## **4.7 Gramática**

El chino pertenece a la familia sino-tibetana, y el español, a la familia indoeuropea. El chino es una lengua radical-aislante, y el español, flexiva. Son dos lenguas muy distintas. Por ejemplo, el chino es una lengua tonal, y el español, no; los caracteres chinos son monosílabos, y las palabras españolas constan casi siempre de dos o más sílabas; en términos generales, el chino no depende de la flexión estricta como

las lenguas de indoeuropeas, sino que principalmente se aprovecha de medios gramaticales como secuencias o partículas para expresar las relaciones y el significado gramatical. No tiene conjugación de verbos, ni tiempos, ni concordancia de género y número...

A continuación, se enumeran algunas características de la gramática de chino.

### 1. Ausencia de flexión.

En comparación con el español, morfológicamente una característica muy destacada del chino consiste en la ausencia de la flexión. En general, las palabras de chino tienen una forma inmutable. No tiene un sistema de cambios flexivos para expresar diferentes significados gramaticales. Es decir, su forma no cambia según el número, género, tiempo, modo, etc. Cuando una palabra se usa en diferentes lugares gramaticales y expresa diferentes significados gramaticales, no hay flexión. Tampoco podemos juzgar la categoría de una palabra desde la morfología si es un nombre, un verbo, un adjetivo o un adverbio, etc. Los casos como la duplicación de palabras no son universales ni obligatorios.

### 2. El cambio de secuencia afecta mucho a la estructura y el significado gramatical.

El orden fundamental de chino es: el sujeto está delante del predicado; el objeto, detrás del verbo; el modificador, delante del modificado; el complemento, detrás de verbo/adjetivo; etc. Cuando la secuencia cambia, como consecuencia, las relaciones estructurales y el significado también pueden cambiar (Zhang Yanqun y Gao Qiliang, 2008).

Pongamos un ejemplo:

他们支持 (ellos apoyan): sujeto+predicado

支持他们 (apoyar a ellos - apoyarles): verbo+objeto

Con el cambio de palabras entre 他们 y 支持, las relaciones no son iguales. El primero es una relación sujeto+predicado, y el segundo, verbo+objeto.

我给他借一本书 (Le pido un libro a él.)。

我借给他一本书 (Le presto un libro a él.)。

Con el cambio de palabras entre 给 y 他, las dos oraciones ya tienen el significado totalmente distinto.

Pero, aquí también vale la pena señalar que en algunos casos, se puede cambiar la secuencia de manera flexible según la necesidad de la expresión. Para la misma relación semántica, hay diferentes formas sintácticas.

他看过了这部电影 (Ha visto esta película)。

这部电影他看过了。

他这部电影看过了。

Estos tres ejemplos utilizan los mismos caracteres. A pesar del cambio de palabras, expresan el mismo sentido: Ha visto esta película. Las relaciones semánticas entre 他 (él), 电影 (película) y 看 (ver) no cambian.

3. El uso de partículas tiene mucha importancia.

Las partículas interpretan un papel importante en la organización de una estructura gramatical. Pueden expresar diferentes relaciones entre las palabras y las locuciones. En algunas estructuras, el uso o no de partículas puede afectar las relaciones gramaticales y la semántica.

Por ejemplo, 更改 (cambiar) 计划 (el plan): verbo+objeto

更改 的 (cambiado) 计 划 (el plan):  
modificador+modificado

Con el uso de “的”, las relaciones gramaticales cambian. El primero es una estructura verbo+objeto, con el sentido de “cambiar el plan”; el segundo, una estructura modificador+modificado, con el sentido de “plan cambiado”.

Otro ejemplo, “牛脾气” significa testarudo, obstinado o terco, mientras “牛的脾气”, el temperamento del toro.

4. No hay relaciones correspondientes fáciles entre las categorías y los elementos sintácticos.

Categoría	Elemento sintáctico	Ejemplos
名词 nombre	主 语 sujeto	电话来了。Ha entrado la llamada.
	宾 语 objeto	打电话 hacer la llamada
	定 语 atributo	电话会议 conferencia de teléfono
	状 语 modificador adverbial	电话联系 contactos por medio de teléfono
	谓 语 predicado	明天星期天。Mañana es domingo.
动词 verbo	谓 语 predicado	他们正在调查这个问题。Están investigando este problema.
	主 语 sujeto	调查研究很重要。La investigación es muy importante.
	定 语 atributo	调查提纲 esquema de investigación
	宾 语 objeto	今天开始调查。Hoy empieza la investigación.
形容词 adjetivo	谓 语 predicado	他的房间很干净。Su habitación está muy limpia.
	定 语 atributo	干净的衣服 ropa limpia
	宾 语 objeto	他爱干净。Él adora la limpieza.
	补 语	洗干净 Lavar limpio

	complemento	
	主 语 sujeto	干净是最要紧的。La limpieza es lo más importante.
	状 语 modificador adverbial	他干净利索地把问题解决了。Ha resuelto el problema de manera pulcra.

Tabla 4.6 Ejemplos de relaciones entre categorías y elementos sintácticos<sup>6</sup>

La Tabla 4.6 muestra ejemplos de relaciones entre categorías y elementos sintácticos. Se nota que hay muchas diferencias entre el español y el chino al respecto. Relativamente, en chino, las relaciones entre las categorías y los elementos sintácticos son más difusas. En español, estas relaciones son más fijas.

Saquemos un ejemplo de la Tabla 4.6.

他的房间很干净。

Su habitación está muy limpia.

En este caso, en español, hay una cópula *ser* o *estar* para servir como predicado. En contraste, en chino, el adjetivo 很干净 solo puede servir como predicado sin la conexión de la cópula.

5. La estructura de palabras, la de locuciones y la de oraciones son parecidas.

De lo pequeño a lo grande, el análisis gramatical de chino tiene cuatro unidades básicas: morfema, palabra, locución y oración. Como resultado de la ausencia de flexión y de la correspondencia entre las categorías y elementos sintácticos, su estructura es básicamente parecida (Zhang Yanqun y Gao Qiliang, 2008).

<sup>6</sup> Oficina Nacional de Propaganda Internacional de chino: “汉语中词类与句法成分之间不是简单的对应关系”. [en línea]: < [http://www.xmuoec.com/gb/teacher/yufa/02/02\\_03\\_2.htm](http://www.xmuoec.com/gb/teacher/yufa/02/02_03_2.htm) > [consulta: 3 de enero de 2011]

El morfema chino es la combinación fonética y semántica más pequeña. La mayoría de los morfemas son monosílabos. Algunos pueden formar una palabra ellos solos, y también pueden formar una palabra junto con otros. Algunos no pueden formar una palabra solos. Si los alumnos pueden dominar la forma, la fonética y el significado de un morfema, les ayudará a entender la semántica de palabras compuestas por este morfema, y al mismo tiempo, les servirá como base para el aprendizaje de locución y oración.

Conocer el principio estructural de la locución puede ayudar a conocer el de la oración. La locución es estática, sin entonación. La oración es dinámica, y lleva una entonación para la oración completa. En chino, cuando agregamos la entonación a muchas locuciones, se pueden convertir en una oración.

Por ejemplo, 洗苹果 (lavar la manzana) es una locución.

En el diálogo “你在干什么？洗苹果。”, “洗苹果” se convierte en una oración.

6. La variedad de clasificadores constituye otra dificultad para muchos alumnos de chino.

Tony McEnery y Richard Xiao ([en línea]) han hecho una investigación de los clasificadores usados en corpus de LCMC y CallHome. Comentan:

Chinese is generally recognised as a ‘classifier language’ not only because of its large inventory of classifiers but also because the use of classifier is mandatory in this language. While it is difficult to give an exact number of classifiers because of the fuzzy boundaries between classifiers and nouns on the one hand and between classifiers and numerals on the other hand (cf. Xing 1993; Li 2000), it has been estimated that there are 500-600 commonly used classifiers in Chinese (cf. Guo 1987: 10). In the LCMC and CallHome corpora used in this study, a total of 421 types of classifiers are found (or 342 types when

reduplicated classifiers and those with a suffix 儿 *er* or 子 *zi* are collapsed into their stems). In terms of tokens, classifiers account for 2.48% of LCMC and 3.13% of CallHome.

There is increasing consensus that classifiers in Chinese can be grouped into three broad categories: nominal (people, animals and objects etc), verbal (actions and events) and temporal (time), but different authors have different ideas of the subcategories for nominal and verbal classifiers. In the taxonomy based on grammatical functions and level of grammaticalization, there are three types of nominal classifiers: specialised (i.e. fully grammaticalized classifiers with no other grammatical function), concurrent (i.e. functioning either as classifiers or some other word classes), and temporary (i.e. other word classes used as classifiers on an ad hoc basis), while verbal classifiers are of two types: specialised (functioning only as classifiers) and borrowed (i.e. temporary borrowings from other word classes) (cf. Guo 1987).

A través de su estudio, se ve claramente que hay gran cantidad y diferentes tipos de clasificadores. Su uso es universal en chino. Además, varios clasificadores pueden juntarse con la misma palabra, y un clasificador puede juntarse con diferentes palabras. Entender las diferencias entre ellos y recordarlos siempre no es un trabajo fácil.

#### 7. (语气词 Interjección 啊, 吗, 吧, etc.)

Hay varias interjecciones como 啊, 吗, 吧, 呢, 呗, etc. Estas palabras se colocan al final de una oración. El significado de la oración puede variar según la interjección empleada.

8. Faltan marcadores explícitos que señalen la relación de los verbos en la “construcción de verbos en serie (*serial verb construction*)” (Comrie, 1990: 825).



Pongamos un ejemplo:

我叫他买橙子吃。

Le digo que compre naranjas para comer.

En chino, aquí no hay marcadores como “*que*” o “*para*”. Las relaciones entre los verbos deben inferirse según el significado y el contexto en que ocurre la combinación de verbos.

Además de lo arriba explicado, el uso de la duplicación, la estructura de complemento resultativo, la oración pasiva sin marcadores, etc, también reflejan las características de la gramática de chino. Para los hispanohablantes que están acostumbrados a la flexión, estos son sus puntos más difíciles.

#### **4.8 Léxico**

La lengua evoluciona sin cesar. Durante este proceso, el léxico también se enriquece.

Para resumir, el léxico chino tiene las características como lo siguiente:

1. La mayoría de los morfemas son monosílabos. Para juzgar la capacidad de componer palabras de un morfema monosílabo, se consideran dos condiciones: uno es si puede formar una palabra solo, el otro es si puede formar una palabra junto con el otro y tener flexibilidad en el orden. Si cumplen ambas condiciones, son morfemas libres. si sólo cumple la segunda, son morfemas semilibres. Los morfemas no libres no cumplen ninguna de las dos. En general, su orden es fijo. No hay muchos morfemas bisílabos y polisílabos. Tampoco tienden a aumentarse (Tian Xiaolin, [en línea]).
2. La composición de palabras y la de locución son parecidas. Eso

se debe a la falta de la flexión. Hay cinco estructuras básicas: yuxtaposición, modificador-cualificativo, verbo-objeto, verbo-complemento y sujeto-predicado. Estas cinco son estructuras principales para el nacimiento de nuevas palabras, sobre todo, las primeras tres (Tian Xiaolin, [en línea]).

3. La mayoría de palabras bisílabas o polisílabas son palabras compuestas. Pero, también hay una parte de palabras bisílabas de un solo morfema, manifestadas principalmente por palabras de duplicación y palabras compuestas de dos caracteres, generalmente aliteradas o rimadas (como 仿佛, 逍遥, 妯娌, etc.). Estos dos tipos de palabras tienen una característica común: en la palabra, el carácter solo no tiene ningún sentido. Sólo la composición de los dos pueden expresar la semántica concreta.
4. Con respecto al extranjerismo, el chino tiene una historia larga de absorber léxico desde otras lenguas. Desde el chino arcaico, ya empezaba a absorber palabras de *Xiongnu* y regiones del Oeste (un término usado en la dinastía *Han* para designar un territorio al oeste de *Yumenguan* que incluye la región actual de Xinjiang y parte de Asia Central (Sun Yizhen, 1999: 887), tales como 琵琶, 葡萄, etc. En el chino contemporáneo, a medida que se intercambia cada vez más con el extranjero, aparecen muchas palabras nuevas que abarcan diversos ámbitos (Shi Youwei, 2000).
5. La tendencia a las palabras bisílabas. Unas palabras monosílabas se transformaban en bisílabas, lo cual constituye una característica importante de la evolución de chino antiguo a chino contemporáneo. Una ventaja radica en que puede evitar el fenómeno de que cuando hay demasiado monosílabas, hay más homófonos y podrá provocar más confusión.<sup>7</sup>

---

<sup>7</sup> Resourced Pool for Teaching and Promoting Chinese as a Foreign Language: “现代汉语词汇的特点” [en línea]: < <http://www.hy123.org/zhishi/word/290.html> > [consulta: 11 de enero de 2011]

6. En comparación con el chino antiguo, los trisílabos han crecido mucho. Especialmente, palabras que llevan sufijos como 性, 化, 式, 员, etc. Aparte, hay gran cantidad de *chengyu*(modismo de cuatro caracteres). Las palabras que llevan más de cuatro ya son términos científicos y especiales.

#### ***4.9 La enseñanza de la lengua china como lengua extranjera en España***

Limitado por el espacio de la tesis, en este apartado, no se va a describir la situación de la enseñanza de chino como segunda lengua en cada país de habla hispana, sino que se proporciona una visión general sobre la enseñanza de chino en España a lo largo de los últimos años, tomándola como una representación, además de resumir los puntos que deberían tenerse en cuenta en la enseñanza de chino hacia los hispanohablantes.

Estos años se ha fortalecido la cooperación política, económica, cultural y educativa entre ambos países y, entre ellas, el intercambio comercial y económico ha crecido notablemente. El volumen total del comercio bilateral para el año 2009 ha sido de 18.364 millones de dólares. Hasta marzo de 2010, España tiene 1707 proyectos de inversión en China<sup>8</sup>.

Bajo esta circunstancia, en la sociedad, cada vez hay más demanda de la gente que habla tanto la lengua castellana como la lengua china, y la enseñanza de chino ha tenido un gran empuje en los últimos años. Según los datos de la Oficina de Asuntos Educativos de la Embajada de la República Popular de China, hasta el final del año 2003, en España sólo había 10 universidades donde se podía estudiar el chino, en las cuales el número de los estudiantes de chino también era muy limitado, puesto que este número no superaba los 1000. En contraste, actualmente, hay más de 40 universidades que han incorporado el

---

<sup>8</sup> Ministerio de Asuntos Exteriores de China: “Las relaciones entre China y España”. [en línea]: [http://www.mfa.gov.cn/chn/gxh/cgb/zcgmzysx/oz/1206\\_41/1206x1/t7362.htm](http://www.mfa.gov.cn/chn/gxh/cgb/zcgmzysx/oz/1206_41/1206x1/t7362.htm) [consulta: 25 de noviembre de 2010]

idioma chino en su oferta académica. Además, la enseñanza del idioma chino está experimentando un crecimiento sin precedente en otros centros docentes ya que más de 100 colegios ya tienen cursos de chino. Un ejemplo de ello nos lo da la Comunitat Valenciana, que ofrece la asignatura del idioma chino como lengua extranjera en unos 30 colegios públicos a partir del año escolar de 2010. En total, actualmente hay más de 25.000 alumnos de chino; y además se han abierto cuatro Institutos Confucio en España, que se encuentran en: Madrid, Valencia, Granada, y Barcelona. Hoy en día, en los cuatro Institutos, el número total de alumnos de chino ha sobrepasado a 2.500.

Este aumento también se ha reflejado en el número de participantes del examen HSK (*Hanyu Shuiping Kaoshi*, examen oficial de nivel de chino). Dicho examen realizó la primera convocatoria en el año 2003 con 13 adscritos, mayoritariamente de origen chino. Mientras que en 2009 se registra el record de 250 adscritos, casi el doble de 2008, y con una participación más activa de españoles que inmigrantes chinos. El primer semestre de 2010, este número ha sobrepasado de 700 personas. Durante todo el año ha sobrepasado los 1.100. Las cifras son elocuentes para demostrar el boom del estudio de chino.

Sin embargo, en comparación con muchos países, la enseñanza de chino en España se inició relativamente tarde, y tiene unas dificultades propias.

La inexistencia de investigaciones sobre problemas específicos para el aprendizaje de la lengua china de hablantes de español, así como el exotismo que hasta hace poco suponía la docencia de la lengua china en España, nos coloca en una situación ciertamente compleja. ...Y a la carencia de docentes formados se une también la enorme escasez de materiales docentes dirigidos a hablantes españoles, adaptados a los diversos objetivos, y que tengan en cuenta los distintos tipos de estudiantes (Fisac, 2008: 40).

En 2009, la Oficina de Asuntos Educativos de la Embajada China en España ha hecho una encuesta a 55 universidades, colegios, escuelas de idiomas e instituciones para preguntarles la situación de la enseñanza de esta lengua.

A continuación, se enumeran algunas dificultades para los hispanohablantes.

### **Pronunciación de ciertas consonantes**

La pronunciación de unas consonantes tales como “<j> ([tɕ]), <q> ([tɕʰ]), <x> ([ç]), <zh> ([tʂ]), <ch> ([tʂʰ]), <sh> ([ʂʰ])” es muy difícil para los hispanohablantes porque no están familiarizados con dichas consonantes, al no estar en los sistemas fonológicos.

### **Manual de chino**

Muchos manuales de chino están escritos y publicados en chino o en inglés; algunos en español están traducidos por personas que no se dedican a la docencia. Así que publicar o encontrar manuales exclusivamente para los hispanohablantes de chino deja mucho que desear puesto que todavía queda mucho por probar e investigar.

### **Diccionarios y gramáticas**

Lo mismo ocurre con diccionarios y gramáticas. Son herramientas útiles y complementarias para el aprendizaje de una lengua extranjera, sin embargo, faltan estos materiales redactados sólo para los hispanohablantes pensando en su característica lingüística y cultural.

### **La metodología**

Debido a la diferencia de mentalidad y cultura entre ambos países y la gran diferencia entre los dos idiomas, hay que ampliar la investigación sobre la metodología de la docencia de la lengua china a los hispanohablantes.

### **El ambiente lingüístico y cultural**

Se recomienda a los alumnos que tengan una estancia en los lugares donde el ambiente lingüístico y cultural es chino, sea por cuenta propia, sea por convenios de intercambio. Con sólo el estudio en clase, sería muy difícil para que los alumnos tengan un alto nivel en la expresión y la comprensión oral.

### **Docentes profesionales**

Con 20 años de esfuerzo, en España ya se han formado cierta cantidad de docentes profesionales. Pero, todavía faltan tanto docentes nativos de chino como docentes hispanohablantes con experiencia y especialidades.

Algunas de estas recomendaciones van a ser tratadas en esta tesis.

---

## **PARTE SEGUNDA**

### **La aplicación**

## **5. Diseño del corpus**

En este capítulo, se exponen los aspectos relacionados con el diseño del corpus, en el que se fundamenta la recogida de grabaciones. Como tarea previa a la realización del corpus en sí, dicho proceso posee esencial importancia para la futura elaboración, explotación de los resultados de la investigación.

### ***5.1 Aspectos generales***

Al hacer el diseño, ante todo, hay que definir bien el objetivo que se persigue con la elaboración del corpus. La creación de corpus puede tener diversos fines, tales como la elaboración de diccionarios, el estudio de frecuencia de palabras, la investigación de algunos fenómenos lingüísticos, la enseñanza de lenguas, la traducción automática, etc. El objetivo decide el tamaño y el contenido de la recolección de datos.

La lengua escrita y la lengua oral se adaptan a distintas circunstancias de uso. En nuestro caso, en vista de que hay muchos corpus a gran escala de la lengua escrita de chino, decidimos elaborar un corpus de habla espontánea para la enseñanza de chino como segunda lengua para hispanohablantes. Un corpus de habla espontánea puede mostrar a los alumnos cómo, de hecho, actúa un nativo en su lengua materna de manera simultánea y dinámica en un contexto situacional, y así ayudarles a desarrollar la competencia comunicativa y expresiva.

El trabajo siguiente será proponer claramente unos determinados factores como tamaño del corpus, tipos de grabación, contenido, participantes, lugares posibles de grabación, formas de recogida de datos, formas de almacenamiento, etc. Todos los pasos son interactivos.

En la elaboración de nuestro corpus hemos preparado un marco que



responde a los siguientes parámetros:

1. **Idioma:** chino *putonghua*. En el capítulo 4, ya explicamos que el chino *putonghua* es la lengua común de los chinos.
2. **Participantes:** universitarios, profesores, empresarios, trabajadores de otras profesiones en lugares públicos, etc. Siempre procedimos a pedirles primero una sencilla declaración firmada, tal como se muestra en el apartado 6.1., de que estaban conformes con la grabación y el uso que se iba a hacer de ella.
3. **Lugares de grabación:** Las grabaciones se han realizado en dos países. Una parte se grabó en la Universidad de Estudios Internacionales de Beijing, y la otra parte, en la Universidad Autónoma de Madrid y Alcalá de Henares. Para un corpus oral, la calidad acústica desempeña un papel muy importante. Si es el caso de que, por una u otra causa, no se escucha bien, por una parte, resultaría difícil para la transcripción. Por ejemplo, el corpus CASS (*Chinese Annotated Spontaneous Speech*) no tenía muy buena calidad acústica y había ruido de fondo, por lo que al final, según los autores, costó mucho trabajo y tiempo hacer la transcripción (Li Aijun et al., 2000). Por otra parte, con estas condiciones sería imposible usarlo en la enseñanza: los alumnos se quejarían del ruido, no entenderían bien el contenido, perderían el interés, etc. El lugar de grabación y los equipos de grabación influyen sin duda mucho en la calidad acústica. Por eso, al hacer la grabación, siempre hemos tratado de buscar un estudio o una oficina silenciosa y usar las grabadoras profesionales para evitar el ruido y cualquier interferencia posible.

El contenido de un corpus tiene que estar sujeto al objetivo para el cual el corpus se construye. Aquí, un aspecto muy importante a tener en cuenta es que los temas deben ser convenientes para nuestro propósito de enseñanza. Una referencia en la que nos hemos basado son los temas que aparecen en *Textos de Español Oral* del laboratorio LLI-UAM. El habla espontánea está enmarcada en un conjunto de circunstancias de carácter social, cultural, espacial, etc. Hemos

pretendido que los temas se relacionen más con la vida cotidiana de los chinos, la cultura china, el aprendizaje de chino, etc.; de ahí que luego el contenido sea interesante y fácil de entender y aparezcan las palabras de uso más cotidiano y frecuente. Los temas seleccionados comprenden aprendizaje, profesión, carácter, ocio, arte, cultura, deporte, viaje, transporte, salud, gastronomía, compras, aficiones, cocina, mascotas, plantas, juego, cosméticos, estaciones del año, etc.

Antes de la grabación, explicamos bien a los participantes el objetivo del corpus y el tema. Todos los archivos de sonido fueron guardados en formato .wav y .mp3 para su futuro procesamiento. En cuanto a la nomenclatura de los archivos de cada grabación, seguimos las convenciones del laboratorio LLI-UAM. Por ejemplo, para el archivo chfo01, aquí *ch* se refiere al idioma chino, *fo*, al subcorpus Formal, y 01, el número de archivo en la carpeta llamada Formal. Chin01 se refiere al archivo número 1 de la grabación de la carpeta de “Informal”. Chme01 se refiere al archivo número 1 de la grabación de la carpeta de “Medios de Comunicación”. La transcripción de cada archivo fue guardada en formato .txt. Cada documento .txt tiene dos partes: la cabecera y la transcripción de la grabación. En la primera parte, vienen todas las informaciones relacionadas con la grabación (cfr. Apartado 6.2), las cuales que pueden ser relevantes en las consultas o el análisis futuro.

Como ya comentamos en el apartado 2.1, en la práctica, no es cierto que cuanto más grande, mejor sea el corpus. El tamaño de un corpus depende del objetivo concreto. “As well as the very large, general corpora designed to assist in writing dictionaries and other reference books, there are thousands of smaller corpora around the world, some comprising only a few thousand words and designed for a particular piece of research (Hunston, 2006: 26).” Para un trabajo como la construcción de un diccionario, como hace falta capturar diferentes aspectos del uso de la palabra, obviamente un corpus pequeño no puede satisfacer tal necesidad. Sin embargo, a pesar de que los corpus generales requieren de mucha inversión de tiempo y recursos y contienen mucha información, hay que decir que no todos los

materiales son convenientes para el estudio de lenguas. Los corpus más pequeños y de ámbito más restringido, que se han creado para una investigación específica o propósito pedagógico, son mucho más adecuados para proporcionar ideas directamente relevantes para la enseñanza y el aprendizaje con propósitos específicos (O'keeffe y McCarthy, 2010).

En nuestro corpus, nos planteamos hacer 10 horas de grabación, equivalente aproximadamente a unas 100.000-120.000 palabras en un corpus español similar.

El número de interlocutores en cada grabación oscila entre el mínimo propio del monólogo y un máximo de tres, ya que un número superior a tres puede aumentar las dificultades de la transcripción y provocar la dificultad de comprensión de los alumnos.

Las fechas de la recogida de los datos se sitúan entre octubre de 2008 y el final del año 2010.

## ***5.2 Elección de los participantes***

Tomando en cuenta que este corpus tiene por objeto formar a los alumnos que quieran estudiar el chino y tengan el español como lengua materna, los materiales obtenidos debían reflejar el habla espontánea de los nativos chinos que han vivido durante largo tiempo en China.

Como es sabido, China es un país de diversas lenguas y dialectos. Aunque la enseñanza de *putonghua* es obligatoria en los colegios, no todos los chinos pueden hablar el *putonghua* estándar. Dependiendo de factores como la región en que se vive, el nivel de educación, el nivel socio-económico, la profesión, etc, el nivel de dominio del *putonghua* varía de persona en persona. En realidad, cierta cantidad de gente habla una variante que es combinación de un *putonghua* estándar y el dialecto de su pueblo (Norman, 2008).

El corpus C-ORAL-CHINA no es un estudio dialectológico. No nos interesa aquí una representación de rasgos lingüísticos diferenciales conforme a la geografía. Al seleccionar a los participantes, es necesario que todos ellos puedan hablar bien –conforme a la gramática estándar– el *putonghua*. Los datos del nivel sociolingüístico de los participantes se han registrado en la cabecera de cada documento. Antes de cada grabación, reiterábamos a los interlocutores la indicación de que hablaran el *putonghua* durante toda la grabación. Si no, creemos que este corpus destinado a la enseñanza de lenguas hubiera perdido las notas de pertinencia lingüística, representatividad y utilidad pragmática.

El corpus C-ORAL-CHINA abarca 10 horas de grabación, y en su obtención han participado 47 miembros. Entre ellos, hay 33 mujeres y 14 hombres. La desigualdad del número de participantes entre mujeres y hombres se debe a varias razones, a saber:

1. La participación en este corpus ha sido voluntaria. Durante la recogida de muestras, primero había que pedir el permiso a cada cual, y el hecho es que hemos tenido más voluntarias que voluntarios como en el caso de corpus español C-ORAL-ROM.
2. La lengua china no es una lengua flexiva. Tampoco hay tantas diferencias entre la lengua femenina y la lengua masculina como el japonés.
3. En el subcorpus de Medios de Comunicación, casi la mayoría de los presentadores de la televisión son mujeres.

Edad	Nº pers.	Porcentaje
A: 18-25 años	35	74.5%
B: 25-40 años	8	17%
C: 40-60 años	1	2.1%
D: >60 años	3	6.4%

Tabla 5.1 Edad de los participantes de C-ORAL-CHINA

La tabla 5.1 nos muestra la estructura de la edad de los participantes. Como tenemos muchos voluntarios universitarios y el subcorpus de Medios de Comunicación pertenece a una universidad china, la edad A (entre 18-25 años) ocupa un porcentaje muy significativo, 74.5%. Una ventaja de esta estructura es que, como son jóvenes, el vocabulario que usa es muy innovador, reflejando el desarrollo de la tecnología y la sociedad.

Ponemos un ejemplo del archivo chin01.

\*SMW: 对 ///α<1138566> 我就是一个 [///] hhh {%act: click} 用一个先进点的词儿说 // 我就是一颜控 ///α<1143226> 就是我干什么都要看外貌的那种 ///α<1146097> 但是也不是说我就是说想找一个特别帅的男朋友这种 // 但我就是喜欢看 ///α<1151854> 就是美女我也喜欢看 // 就是好看的东西我都看 ///α<1155521>

“XX 控” es una forma muy nueva de expresar la obsesión de una persona por algo favorito suyo. El uso de este tipo de palabras (neologismos) representa las palabras recién aparecidas en aquella misma época en que hacemos la recogida de datos.

En la transcripción, también hemos notado un fenómeno. En la grabación, mucha gente mantiene sus acostumbradas muletillas, pero casi nadie dice palabrotas. Creemos que eso se relaciona con el hecho de que ya saben cuál es el uso final de este corpus y la grabación no es secreta, lo cual afecta a su comportamiento lingüístico. Para un chino, pronunciar tacos en una ocasión así, es una falta de cortesía.

Además, muchos participantes de nuestro corpus tienen experiencia de haber vivido cierto tiempo en países hispanohablantes. En unos documentos, han contado su vida en esos países, lo cual puede ser interesante e atractivo para los alumnos hispanohablantes y les facilita la comparación cultural entre distintos países.

### ***5.3 Tres tipos de grabación***

Diferentes corpus adoptan diferentes estrategias de estructura. En C-ORAL-CHINA, usamos la partición de C-ORAL-ROM. Es decir, dividimos el habla en formal (60%) e informal (40%). El habla formal se divide en dos partes: medios de comunicación y el habla en contexto público.

Según Massimo Moneglia (2005), si bien es cierto que un registro formal de expresión con frecuencia se basa parcialmente en un texto escrito, también está ligado a otras condiciones relevantes, tales como:

- a. uso público del habla;
- b. uso profesional del habla, de acuerdo con la función social y el papel del hablante en la comunidad;
- c. la tarea deliberada de llevar a cabo un texto oral que implica el tratamiento de un tema, argumentación, conclusiones, etc.

Concluye este autor diciendo que el habla formal todavía puede ser espontánea siempre y cuando no sea la ejecución de un texto previamente preparado y escrito.

Distribuimos las 10 horas en tres subcorpus: Formal en contexto público, Medios de comunicación e Informal. La duración y el contenido de cada subcorpus se diseñan como la tabla 5.2.

<b>Tipos</b>	<b>Duración</b>	<b>Contenido</b>	<b>Lugares de grabación</b>
Formal en contexto público	2 horas	Conferencias sobre la cultura china, clases de chino, etc.	China y España
Medios de comunicación (radio y televisión)	4 horas	Entrevistas, reportajes, etc.	China
Informal	4 horas	Diálogos en sitios públicos, diálogos en sitios privados	China y España

Tabla 5.2 Parámetros en el diseño del corpus C-ORAL-CHINA

Entre ellos, en el subcorpus Formal en contexto público, todos los textos orales son monólogos. Según Patrice Pavis, el monólogo puede ser caracterizado como el discurso de un personaje que no está dirigido directamente a un interlocutor con el propósito de obtener una respuesta (Patrice Pavis, 1998). En ese caso, el emisor emite las informaciones a los receptores para convencerles de lo que está contando.

El subcorpus Informal abarca 25 diálogos. El diálogo es un tipo de interacción entre dos interlocutores. Es una forma de conversación más cooperativa. En comparación con el monólogo, Briz (2002) lo caracteriza por los siguientes rasgos:

- a) Dialógico, lo que implica sucesión de intercambios;
- b) Inmediato, puesto que, a diferencia de un informativo o un mensaje pregrabado, se desarrolla en la coordenada espacio-temporal aquí-ahora-ante ti;
- c) Retroalimentado y cooperativo, puesto que se obra juntamente con otro y su intervención;

d) Dinámico, como demuestra la alternancia de turnos, que además es no predeterminada.

Aunque el tema es definido antes de hacer la grabación, el contenido no es planificado. En realidad, no hay control del contenido ni del tiempo en la producción del habla por parte de la investigadora. Por lo tanto, hay parejas que hablan mucho tiempo y parejas que hablan apenas un par de minutos.

#### ***5.4 Comparación con el corpus CADCC***

Como se expone en el capítulo 2, actualmente hay diferentes corpus de habla espontánea en China. Aquí vale la pena mencionar los corpus orales explotados por la Academia de Ciencias Sociales de China (*Chinese Academy of Social Sciences*):

- Corpus de Monosílabos. Contiene 1275 monosílabos, pronunciados por 15 hablantes masculinos.
- Corpus de Bisílabos. Abarca más de 3000 combinaciones de vocales y consonantes.
- *Annotated Speech Corpus of Chinese Discourse* (ASCCD). Este corpus leído está formado por textos orales, datos fonéticos y la anotación. Se puede usar en la investigación fonética, explotación de ingeniería fonética y la enseñanza de *putonghua*.
- *Chinese Annotated Dialogue and Conversation Corpus* (CADCC). Se compone de grabación de diálogos de lengua natural y la transcripción.
- *Chinese Annotated Spontaneous Speech* (CASS)  
Corpus de Reconocimiento de Habla “863” / Corpus de Síntesis de Habla “863”.

En este apartado, seleccionamos uno famoso de entre ellos CADCC (*Chinese Annotated Dialogue and Conversation Corpus*) para hacer una comparación de diferentes aspectos entre los dos.



CADCC está destinado a la investigación de la lengua natural, al reconocimiento fonético y a la enseñanza de *putonghua* de nivel alto, etc. El sonido fuente ha sido grabado en formato .wav, 16 bits, 16 KHZ, con un almacenamiento de 1.6 GB (Liu Yabin y Li Aijun, [en línea]).

#### 1. La estructura.

CADCC abarca dos subcorpus: Set1 (de diálogos telefónicos) y Set2 (de diálogos en conversación). La Tabla 5.3 proporciona algunos datos básicos de este corpus. Tiene 12 unidades de diálogos, captados en un ambiente de grabación profesional. En cada unidad participan dos interlocutores. Todos hablan el estándar *putonghua*.

	Set1	Set2
<b>Contenido</b>	Servicio en hotel	Sin restricción
<b>Duración</b>	2 horas	16.2 horas
<b>Participante</b>	>200	16 hombres y 10 mujeres
<b>Acentos dialectales</b>	Sí	no
<b>Transcripción</b>	sí	sí
<b>Formato</b>	.wav	.wav
<b>Solapamiento</b>	más	menos

Tabla 5. 3 Datos básicos del corpus CADCC

CADCC no tiene ni monólogos ni grabaciones en medios de comunicación como el nuestro. En este sentido, C-ORAL-CHINA abre a más variedades de habla espontánea.

#### 2. Diseño de temas

En CADCC, no han trazado ningún límite en el contenido del diálogo. Este corpus pretendía reflejar absolutamente las características de la lengua natural en un contexto real. Se trata de conversaciones libres, sin control alguno por parte del investigador.

En C-ORAL-CHINA, la conversación es semidirigida, con control del investigador. Es decir, antes de hacer la grabación, se les explica a todos los participantes el propósito específico del corpus y se les propone un tema preferido previamente escogido.

El objetivo del diseño puede explicar dicha diferencia. En CADCC se trataba de hacer un estudio de tiempo/tema.

<b>Hablante</b>	<b>Sexo</b>	<b>Número total de tópicos</b>	<b>Duración en seg.</b>	<b>Seg. por tópico</b>
SUNXI	MEN	42	6524.859	155.3538
ZHNGJ	MEN	35	4197.666	119.9333
XINGY	WOMEN	14	1953.768	139.5549
CHENX	WOMEN	13	3062.002	235.5386
DUYU	WOMEN	27	4002.587	148.244
SONGW	MEN	11	2911.886	264.7169
LIUJI	MEN	10	2718.652	271.8652
LVJIN	WOMEN	27	4177.172	154.7101
XUCHA	MEN	15	6221.138	414.7425
DURUI	MEN	32	5859.053	183.0954
TANJI	MEN	7	2531.122	361.5889

Tabla 5.4 Distribución de tiempo/tema de cada pareja de interlocutores

En el Set2 de CADCC, cada pareja son colegas o compañeros que tienen la misma afición o temas comunes. El diálogo dura más o menos una hora. La tabla 5.4 trata de la distribución de tiempo/tema de cada pareja de interlocutores. Según ellos, 80% de los temas duran menos de 4-5 minutos, y 90% de los temas, menos de 5-6 minutos.

En C-ORAL-CHINA, como el tema es fijo, se nota que algunas

parejas son capaces de hablar más tiempo permaneciendo exclusivamente en el ámbito de la cuestión propuesta. No es el caso de otras que, eventualmente, dentro de una misma grabación pueden desviarse a otros temas. En cualquier caso, unas y otras dedican lo esencial de la conversación al tema de que se trata y que se indica en el encabezamiento del texto de cada grabación. Nuestra idea fundamental con esto es que, luego, aprovechando nuestro corpus, los alumnos de chino puedan familiarizarse con ciertas palabras habituales y sinónimos de un mismo tema o campo semántico.

### 3. La grabación secreta y los participantes.

En CADCC, la grabación se hace en oficinas o habitaciones normales. Los equipos de grabación inalámbricos se colocan en otra habitación. Cada interlocutor lleva micrófono inalámbrico encima. Lo han hecho así para poder garantizar que ambas partes puedan entrar en el diálogo de manera más natural.

En C-ORAL-CHINA, la grabadora no se esconde. Las interacciones se han grabado con el micrófono a la vista, y se han realizado casi en su totalidad frente a la investigadora. La presencia de la investigadora sirve para garantizar que todos los equipos funcionen bien y la grabación tenga buena calidad.

### 4. El estudio de la velocidad al hablar.

En CADCC, aparte de los dos subcorpus, grabaron a 6 locutores que leían un artículo seleccionado. Después, hacían una comparación entre CADCC, el corpus leído ASCCD y estas grabaciones de texto leído. En nuestro corpus, no incluimos el subcorpus leído. Teniendo en consideración que actualmente la mayoría de los manuales lleva un CD para practicar la lectura, no lo hemos preparado.

Según los investigadores del otro estudio, en su experimento, la velocidad de la lengua espontánea es cuatro veces mayor que la velocidad del corpus leído. Por nuestra parte, también hemos realizado una medición de velocidad en los contenidos del corpus.

<b>Archivo</b>	<b>Segundos</b>	<b>Nº Caracteres</b>	<b>Caracteres / Segundo</b>
chfo01	1334	4349	3.26
chfo02	925	3663	3.96
chfo03	611	2096	3.43
chfo04	368	1358	3.69
chfo05	148	530	3.58
chfo06	465	1392	2.99
chfo07	218	745	3.42
chfo08	286	1011	3.53
chfo09	386	1336	3.46
chfo10	295	993	3.37
chfo11	304	1060	3.49
chfo12	268	955	3.56
chfo13	92	283	3.08
chfo14	80	238	2.98
chfo15	388	1247	3.21
chfo16	84	241	2.87
chfo17	105	313	2.98
chfo18	244	738	3.02
chfo19	133	396	2.98
chfo20	446	1472	3.30
chfo21	125	414	3.31
Media			3.4

Tabla 5.5 Velocidad caracteres / segundo del subcorpus Formal en contexto público.

Archivo	Segundos	Nº Caracteres	Caracteres / Segundo
chin01	2076	9291	4.48
chin02	428	1978	4.62
chin03	283	1254	4.43
chin04	603	2896	4.80
chin05	54	232	4.30
chin06	360	1779	4.94
chin07	335	1462	4.36
chin08	810	2924	3.61
chin09	964	4632	4.80
chin10	73	258	3.53
chin11	206	710	3.45
chin12	428	2018	4.71
chin13	980	4594	4.69
chin14	401	1260	3.14
chin15	327	1281	3.92
chin16	108	373	3.45
chin17	528	2301	4.36
chin18	99	444	4.48
Chin19	954	3690	3.87
chin20	300	1026	3.42
chin21	101	342	3.39
chin22	869	4448	5.12
chin23	1514	5610	3.71
chin24	576	2638	4.58
Chin25	483	2056	4.26
Media			4.29

Tabla 5.6 Velocidad caracteres / segundo del subcorpus Informal

Archivo	Segundos	Nº Caracteres	Caracteres / Segundo
chme01	480	906	1.89
chme02	651	2015	3.10
chme03	924	2860	3.10
chme04	692	1947	2.81
chme05	47	172	3.66
chme06	259	660	2.55
chme07	581	2735	4.71
chme08	603	2409	4.00
chme09	1006	3648	3.63
chme10	63	272	4.32
chme11	102	390	3.82
chme12	96	304	3.17
chme13	728	2834	3.89
chme14	635	2562	4.03
chme15	940	3920	4.17
chme16	1839	7335	3.99
chme17	377	1368	3.63
chme18	871	3278	3.76
chme19	83	270	3.25
chme20	540	1924	3.56
chme21	818	2851	3.49
chme22	291	949	3.26
chme23	527	2570	4.88
chme24	1089	4725	4.34
chme25	840	3472	4.13
Media			3.74

Tabla 5.7 Velocidad caracteres / segundo del subcorpus de Medios de comunicación

Las tres tablas 5.5, 5.6 y 5.7 presentan la velocidad de habla calculada en cada texto oral y la velocidad media en los tres subcorpus. Nos damos cuenta de que la rapidez al hablar en el subcorpus Informal es mayor que en la de los dos subcorpus Formal.

#### 4. Convenciones de transcripción.

En CADCC, la transcripción también se almacena en formato .txt. La transcripción de cada enunciado tiene tres líneas de información. La figura 5.8 es un ejemplo extraído.

```

1: (67140:28306) B:[NS].
Word Segments: \B:[NS].\
Pinyin String: B:[NS].

2: (95446:73585) A:[BR]哎呀[MO][SG][SM].
Word Segments: \A:[BR]\哎呀\[MO][SG][SM].\
Pinyin String: A:[BR] ai1 ya1 [MO][SG][SM].

3: (169031:24955) B:你们现在[DS]紧不紧呀[MO]?
Word Segments: \B:\你们\现在\[DS]\紧\不\紧\呀\[MO]?
Pinyin String: B: ni3 men0 xian4 zai4 [DS] jin3 bu4 jin3 ya0
[MO]?

4: (193986:46226) A:[BR]紧不紧就看自己了,我这两天还行.
Word Segments: \A:[BR]\紧\不\紧\就\看\自己\了\,\我\这两\天\还\行\.\
Pinyin String: A:[BR] jin3 bu4 jin3 jiu4 kan4 zi4 ji3 le0 ,
wo3 zhei4 liang3 tian1 hai2 xing2 .

5: (240212:8828) B:全自己安排.
Word Segments: \B:\全\自己\安排\.\
Pinyin String: B: quan2 zi4 ji3 an1 pai2 .

```

Figura 5.8 Ejemplo de transcripción en CADCC

Entre ellas, en la primera línea, el primer número indica el número del enunciado; entre comillas, dos cifras se refieren respectivamente al inicio temporal de este segmento de sonido y su duración.

La segunda línea es la información de segmentación de palabras.

La tercera línea es la transcripción en *pinyin*.

En C-ORAL-CHINA, considerando que el nivel de chino de cada alumno es distinto y que en realidad los chinos leen los caracteres y no el *pinyin*, no metemos las dos transcripciones en el mismo documento. Lo que sale en la transcripción de la página web es como la figura 5.9:

1	0-2.785	CRJ: 同学们好 !
2	2.785-5.232	CRJ: 今天是我们上课的第一天 ///
3	5.232-9.907	CRJ: 我先来介绍一下我们这门课 // 怎么 hhh {%com: close the door} 来上 ///
4	9.907-15.606	CRJ: 这一门课我们学习的是 // 新实用汉语课本的第二册 ///
5	15.606-21.853	CRJ: 这个学期 / 我们学习 / 第十五课 / hhh {%com: open the book} 到第二十课 ///
6	21.853-24.33	CRJ: 一共我们学习六课书 ///
7	24.33-28.957	CRJ: &mm 每一课书基本上是用六个小时 ///
8	28.957-38.243	CRJ: 前 [/] 前四个小时或者五个小时 // 我们学习 / 生词 / 语法 / 课文和一部分的练习 ///

Figura 5.9 Ejemplo de transcripción en C-ORAL-CHINA

Igual que en CADCC, el primer número indica el número del enunciado. Las dos cifras que lo siguen son el tiempo inicial y terminal de este segmento.

Preparamos otro documento .txt con la transcripción de *pinyin*.

Presentamos dos formas de *pinyin*.



\*CRJ: tóng xué men hǎo! jīn tiān shì wǒ men shàng kè de dì yī tiān。 wǒ xiān lái jiē shao yí xià wǒ men zhè mén kè, zěn me hhh {%act: close the door} lái shàng。 Zhè yì mén kè wǒ men xué xí de shì, xīn shí yòng hàn yǔ kè běn de dì èr cè。 zhè ge xué qī wǒ men xué xí dì shí wǔ kè hhh {%act: open the book} dào dì èr shí kè。 yí gòng wǒ men xué xí liù kè shū。 ng..... měi yí kè shū jī běn shàng shì yòng liù ge xiǎo shí。 qián ..... qián sì ge xiǎo shí huò zhě wǔ ge xiǎo shí, wǒ men xué xí shēng cí、 yǔ fǎ、 kè wén hé yí bù fēn de liàn xí。

Figura 5.10 Ejemplo de transcripción de *pinyin* (sin segmentación de palabras)

\*CRJ: tóngxuémén hǎo! jīntiān shì wǒmen shàngkè de dìyītiān。 wǒ xiānlái jièshao yíxià wǒmen zhè mén kè, zěnmehhh {%com: close the door} lái shàng。 zhè yì mén kè wǒmen xuéxí de shì, xīn shíyòng hàn yǔ kè běn de dì èr cè。 zhè ge xuéqī wǒmen xuéxí dì shí wǔ kè hhh {%com: open the book} dào dì èr shí kè。 yí gòng wǒmen xuéxí liù kè shū。 ng..... měi yí kè shū jī běn shàng shì yòng liù ge xiǎo shí。 qián ..... qián sì ge xiǎo shí huò zhě wǔ ge xiǎo shí, wǒmen xuéxí shēngcí、 yǔfǎ、 kèwén hé yí bù fēn de liàn xí。

Figura 5.11 Ejemplo de transcripción de *pinyin* (con segmentación de palabras)

En las dos figuras 5.10 y 5.11, la primera forma no dispone de la segmentación de palabras, mientras la segunda dispone de ella. De esta manera, en la búsqueda, los alumnos podrán buscar por carácter o por *pinyin* como quieran. Si alguno no sabe el principio de la segmentación de palabras, también puede usar la herramienta para encontrar la palabra.

Aparte de todo lo dicho, otra diferencia de la transcripción de *pinyin* es que en CADCC, en vez del estándar *pinyin*, usan solamente las letras del abecedario más el número, sin ningún símbolo diacrítico. Pero esto resulta, a nuestro juicio algo tosco y poco operativo. Además, una transcripción así puede ser difícil de leer y, sobre todo, no se corresponde a la costumbre de los chinos. Por eso, hemos querido que nuestra transcripción fuera una estandarizada y usada realmente por los chinos. Gracias a las codificaciones modernas como UTF8 de Unicode, en C-ORAL-CHINA, este problema está resuelto. Es más, hemos notado que nuestra transcripción no afecta el

estudio de los fonemas (véase el apartado 7.2.).

En cuanto a la transcripción de los llamados signos o fenómenos paralingüísticos, nuestro método (C-ORAL-CHINA) y el de CADCC son también distintos.

	<b>Fenómeno</b>		<b>Transcripción</b>
1	拖音	lengthening	[LE ]
2	喘息	breathing	[BR ]
3	笑声	laughing	[LA ]
4	哭声	crying	[CR ]
5	咳嗽	coughing	[CO ]
6	间断	disfluency	[DS ]
7	噪音	noise	[NS ]
8	沉默	silence (long)	[SI ]
9	含混音	murmur/uncertain segment	[UC ]
10	语气词	modal/exclamation	[MO ]
11	咂嘴音	smack	[SM ]
12	其他语言	non-Chinese	[NC ]
13	吸鼻音	sniffle	[SN ]
14	打哈欠	yawn	[YA ]
15	重叠发音	overlap	[OV ]
16	插话	interjection	[IN ]
17	吞咽	deglutition	[DE ]
18	清嗓子	hawk	[HA ]
19	打喷嚏	sneezes	[SE ]
20	电话忙音	beep	[BP ]
21	颤音	trill	[TR ]
22	口误	error	[ER ]
23	填充停顿	filled pause	[FP ]

Tabla 5.12 Convenciones de transcripción de signos paralingüísticos de CADCC

En CADCC, usan la abreviatura preelaborada para transcribir los signos paralingüísticos como la tabla 5.12.

Por ejemplo:

A:LE<MO<噢 LE> MO>;B:那个就是 DS<一一 DS>锤子的事儿。

En el nuestro, no usamos la abreviatura, sino una descripción más explícita para que el lector se aclare con el fenómeno de un vistazo (esto se concreta en el apartado 6.2.2.6).

Ejemplo:

\*WAL: &mm 我想西班牙人 [///] &mm 我想很 [/] 对于很多西班牙的同学来讲 // 那个 hhh {%act: clearing of the throat} 上这个课的目的呢 / 肯定是想将来 / 有机会能够自己 / 比如在中国呀或者在西班牙也好 // 能够 hhh {%act: click} &mm 上一些新闻的看 [/] 上一些新闻网站啊 / 看一些实时的国内一些消息 ///

### **5.5 Diseño final**

El corpus C-ORAL-CHINA consta de 140.703 caracteres. En total son diez horas de grabación, distribuidas en tres subcorpus, con 71 textos, con la participación de 47 voluntarios.

Todos los textos orales se ofrecen en formato electrónico digital. Representan el estándar *putonghua*. Incluyen diferentes situaciones comunicativas en la vida real. Además, los alumnos pueden consultarlo a través de la página web.

En el Apéndice, se adjuntan ejemplos de transcripción y las tablas de las informaciones generales de los subcorpus sobre el número de palabras, duración de los archivos, números de turnos, enunciados por fichero, etc.

## 6. Metodología

### *6.1 Grabación y digitalización*

Las mayor parte de las grabaciones se han realizado con una grabadora Edirol (modelo R09HR) con dos micrófonos incorporados. El sonido fuente ha sido grabado en formato WAV, 16 bit, 44100 Hz, estéreo.

La parte de medios de comunicación del corpus ha sido proporcionada directamente por la Radio y la Televisión de la Universidad de Estudios Internacionales de Beijing.

La protección de los derechos de propiedad intelectual y de privacidad es obligatorio en los corpus actuales. Los participantes en la grabación tienen derecho a prohibir que se publique o se comercialice su voz. Antes de hacer la grabación, explicamos a todos los participantes el uso final de su grabación y les pedimos que firmaran un permiso en el que nos autorizaban la grabación y el uso de la voz. En la Figura 6.1, presentamos el modelo de permiso que usamos en el C-ORAL-CHINA siguiendo la convención del laboratorio LLI-UAM.

I, ....., at the request of the person in charge of the project carried out by the Computational Linguistics Laboratory (LLI) at the Universidad Autónoma de Madrid, whose purpose is to develop a spontaneous speech corpus made up of a collection of recordings and transcriptions, within the framework of the Doctoral Dissertation by ....., give my permission and authorization for the following:

1. Recording of my voice.
2. Transcription of the recording.
3. Treatment of the sound and the transcription.
4. Dissemination and commercialization of the sound and the transcription
5. I keep the right to listen to the recordings and to deny my authorization for whatever reason that I consider pertinent.

The persons in charge guarantee the anonymity of the participants and of any person mentioned in the recording.

Date:

Signature:

Figura 6.1 Modelo de permiso de grabación

Durante todo el proceso, había que observar si funcionaban bien la grabadora y el micrófono. En la transcripción, si se descubría que había alguna grabación o algún trozo de grabación que no se podía escuchar bien, lo quitábamos o lo cortábamos.

En contadas ocasiones se ha empleado el programa Cool Edit para manipular el sonido mejorando su calidad, eliminando ruidos o cortando partes que no son relevantes o que preferimos no utilizar por cuestiones de anonimato e intimidad. De esta forma seleccionamos las grabaciones que vamos a transcribir, editándolas en archivos .wav independientes.

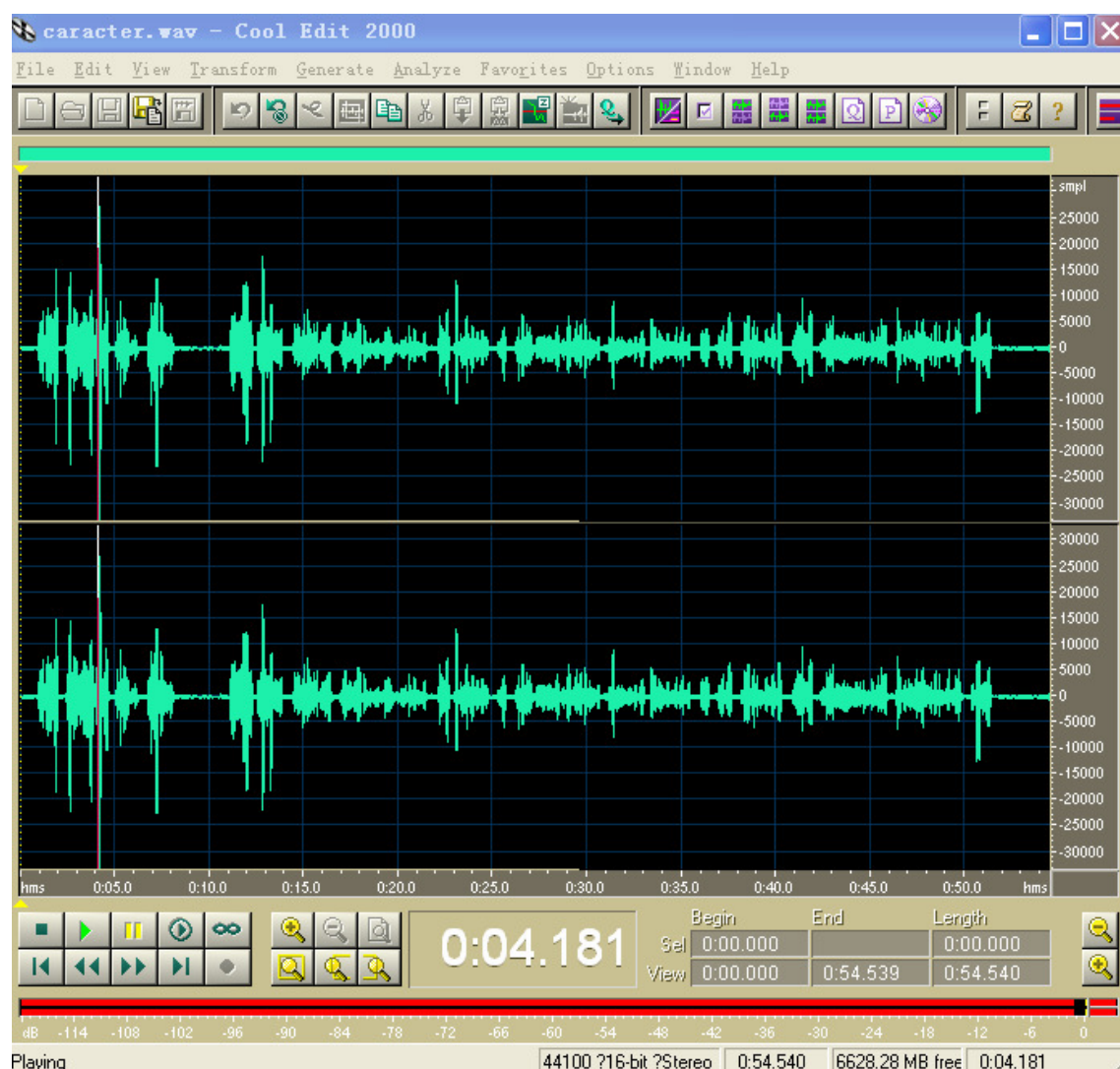


Figura 6.2 Edición del sonido con Cool Edit

## **6.2 Transcripción ortográfica**

En comparación con el corpus de lengua escrita, la construcción del corpus de habla espontánea es más complicada y difícil. El habla espontánea contiene muchos fenómenos como pausas, palabras cortadas, vacilaciones, repeticiones, etc. La transcripción de este tipo de corpus es muy costosa. Los programas actuales de reconocimiento de habla no son capaces de hacer la transcripción ortográfica adecuada automáticamente. Tenemos que hacerlo manualmente. En nuestro caso, una hora de grabación puede costar 40 horas de trabajo.

La transcripción constituye el punto de partida para el procesamiento posterior del corpus y los análisis lingüísticos. En C-ORAL-CHINA, se procura atenerse a 3 principios.

1. La fidelidad. Tenemos que ser fieles a los enunciados, manteniendo el estado original sin hacer ninguna corrección u omisión.
2. La integridad. Durante el habla oral, debido a varios motivos, es inevitable que haya oraciones incompletas, pausas, interrupciones, autocorrección, etc. Aunque algunos elementos no tienen una representación gráfica normativa, siempre tratamos de hacer la transcripción de manera integral elaborando algunas convenciones para registrarlos porque todo esto impulsa la marcha de la comunicación del lenguaje.
3. La exactitud. Se refiere a juzgar con exactitud el contenido de la grabación como la pronunciación, la entonación, las locuciones y el léxico, etc, usando correctamente los signos de transcripción ya establecidos.

De acuerdo con estos principios, seguimos las convenciones del laboratorio LLI-UAM, en combinación con las características reales de la lengua china. Sin duda, un estándar favorece tanto la buena marcha de la transcripción como los análisis siguientes a la transcripción.

Cabe señalar finalmente que para garantizar la precisión de la transcripción del sonido y los fenómenos paralingüísticos, es imprescindible la revisión por un transcriptor diferente.

El procedimiento de transcripción ortográfica se divide en tres fases:

1. Transcripción ortográfica con inclusión simultánea del etiquetado de marcas prosódicas.
2. Revisión del etiquetado por un segundo etiquetador.
3. Segunda revisión por parte del primer etiquetador para garantizar al máximo la precisión.

El texto de una transcripción se va a estructurar en dos partes: cabecera<sup>9</sup> y la propia transcripción del sonido. Así pues, el formato textual comprende dos niveles:

1. **Metadatos:** información de la sesión de grabación (contexto, tamaño, participantes, calidad acústica, fuente, transcriptores y revisores). Estos datos significan mucho para la investigación lingüística: A través de ellos, se puede consultar todo tipo de informaciones relacionadas con el corpus; se puede formar distintos subcorpus para adaptarse a diferentes necesidades de los investigadores; se puede hacer comparaciones entre los subcorpus para estudiar y descubrir los factores que puedan influir en el uso y del desarrollo de la lengua, etc.
2. **Representación dialógica:** información lingüística y no lingüística presentada en forma textual y con dos niveles de información:

- Líneas de texto: transcripción ortográfica.
- Líneas dependientes: información contextual.

---

<sup>9</sup> Para la información contenida en la cabecera se utiliza el inglés por ser la convención seguida en el proyecto C-ORAL-ROM.



### 6.2.1 Cabeceras (Metadatos)

Aquí, se presenta toda la información de la grabación.

@Title: título que ayude a reconocer el texto.

@File: nombre del archivo.

@Participants: datos de cada participante, donde se incluye:

- Nombre: tres letras mayúsculas, generalmente las tres primeras letras del nombre del participante.
- Sexo (man/woman).
- Edad: se utiliza el código A: 18-25; B 25-40; C: 40-60; D: >60.
- Educación: 1: primaria o menos; 2: secundaria; 3: estudios superiores.
- Profesión.
- Role: papel del participante en la interacción.

@Date: separado por barras, con las cuatro cifras del año.

@Place: lugar geográfico de la grabación.

@Situation: breve descripción del espacio e información relevante del lugar en el que se ha hecho la grabación.

@Topic: temas y subtemas tratados en el texto, con un máximo de 50 caracteres.

@Source: en este caso indicamos el nombre del proyecto.

@Class: tipo de interacción, es decir, formal, informal, privada, pública, etc.

@Length: minutos (') y segundos (").

@Characters: número de caracteres.

@Acoustic\_quality: A (muy buena), B (buena), C (suficientemente buena).

@Transcriber: nombre del transcriptor.

@Revisor: nombre del revisor.

@Comments: comentarios del transcriptor sobre el texto.

Pongamos un ejemplo:

@Title: claseperiodico  
@File: chfo02  
@Participants: WAL, (man, B, 3, profesor, participant)  
@Date: 28/10/2009  
@Place: Madrid  
@Situation: in a classroom, not hidden  
@Topic: newspaper, spanish students, learning chinese  
@Source: C\_ORAL\_CHINA  
@Class: formal, public, speech  
@Length: 15' 25''  
@Characters: 3740  
@Acoustic\_quality: A  
@Transcriber: Dong Yang  
@Revisor: AZY  
@Comments:

## ***6.2.2 Convenciones***

### ***6.2.2.1 Turnos***

\*ABC: los turnos se indican con asterisco, tres mayúsculas (pueden ser las tres primeras letras del nombre del participante. Si el nombre sólo tiene dos caracteres, pueden ser las dos primeras letras del apellido más la primera letra del nombre), dos puntos y un espacio.

\*JIJ: 听说你会下象棋 ? 那么 // 是国际象棋 / 还是中国象棋 ?

\*HXL: 我会 [/] 我会下国际象棋啊 ///

### ***6.2.2.2 Comentarios***

En los comentarios se introduce toda la información paralela a la conversación que se considera relevante para comprender lo que está

sucedendo.

Los comentarios se escriben dentro del turno, en la misma línea, entre llaves { }, inmediatamente después de la palabra tras la cual se ha producido el fenómeno.

Como se podrá observar, los comentarios se escriben en inglés. En función de la clase de información que se quiera expresar en el texto, existen distintos tipos de comentarios:

- **%act:** son los comentarios más usuales y sirven fundamentalmente para expresar el paralenguaje:

\*JII: 那一个五岁的小孩儿就下国际象棋 // 会不会让人有那种 // 幼年老成的感觉呢 hhh {%act: laugh} ?

- **%com:** sirve para hacer referencia a algún cambio ocurrido en la situación, como por ejemplo, que de pronto un participante se levante, se cambie de lugar, o se marche de la escena, o también la llegada de otro participante. Todo lo que pueda influir en la forma de enfocar la situación comunicativa y en los temas de conversación. Por ejemplo, encender la televisión y comentar los anuncios publicitarios.

\*CRJ: hhh {%com: open the book} 我再讲一下这本书 [/] 它的这本书的 / 这个 // 主要的安排的 / 想法 // 就是它的主要的想法 ///

Por otra parte, en este corpus, hay cuatro horas de grabaciones de medios de comunicación. Entre ellas, algunas tienen la música como fondo para unos trozos o durante toda la grabación.

Se usa {%com: background music} para indicar el inicio de la música.

\*WYL: {%com: background music} Hello / 观众朋友们大家好 !

Si la música sale en toda grabación, este comentario sólo aparece una vez.

Si la música sale en algún trozo, cuando la música termina, se usa `{%com: end of background music}`. En el trozo siguiente, los comentarios se repiten al inicio y al final de la música.

\*WYL: 还有 // 良好的生活习惯 / 对皮肤的好坏可以起到关键性作用哦 `{%com: end of background music}` !

Lo mismo pasa con otro tipo de sonidos.

\*WYL: `{%com: background music and street noise}` 看了那么多五颜六色 // 精美绝伦的镜头 // 是不是给你的眼球儿 / 带来了不小的享受和冲击呢 ? 现在的你是不是已经有了灵感 ? 如何把你的小窝儿装点一新了呢 ? 刚才 // 我们给大家介绍了那么多有关如何布置宿舍方面的知识和小窍门 // 希望能对大家有所帮助 ! 好了 // 节日宿舍布置好了 // 下面就是挑选什么礼物送给朋友 / 会别有新意 / 不落俗套 / 而又经济实惠呢 ? 下面就跟着我一起去挑选节日礼物吧 `{%com: end of background music and street noise}` !

Si en un enunciado, sólo hay música, se transcribe como `{%com: music}`.

\*WYL: 比如像帽子 // 围巾 // 手提包 // 还有靴子 /// `{%com: music}` /// 一件合适的羽绒服 // 在今年冬天已经成为了一件热门的时尚单品 ///

Siempre se escribe dentro del turno, en la misma línea, entre llaves `{ }`, justamente cuando suena o para el sonido de fondo.

• **%alt**: este tipo de comentario tiene dos funciones: en primer lugar, se utiliza para reflejar los errores de producción. Por ejemplo, imaginemos que un hablante en un momento determinado de la

conversación se equivoca y dice “pacharrón” en lugar de “chaparrón”. En nuestro caso, en la transcripción se presentaría así:

\*LIC: 啊 / 说到排球 / 我也蛮感兴趣的 /// 您们 {%alt: 您几位} 一般都什么时间去打呀 ?

Así pues, en la transcripción se refleja lo que el hablante dijo realmente, y la forma normativa se pone en el comentario. En segundo lugar, los comentarios %alt sirven para recoger todo tipo de expresiones o fenómenos que son típicos del habla oral pero que no están aceptados por la norma. Por ejemplo, imaginemos que un hablante dice “fúbol” en lugar de “fútbol” o “m’acuerdo” en lugar de “me acuerdo”. En todos estos casos, la forma que, desde un punto de vista normativo, se considera correcta se pone en el comentario de la transcripción, mientras que la variante lingüística se pone en el texto.

- **%pho:** este comentario se ha utilizado para reproducir las pronunciaciones erróneas (en la misma línea entre llaves).

\*XYZ: &mm 圣 {%pho: [sèng]} 诞节吗 // &mm 因为我们  
是 {%pho: [sì]} 马上 {%pho: [sàng]} 要毕业了 // 可  
能会跟全班同学一起聚一聚 // 吃 {%pho: [cī]} 顿饭 //  
然后会去唱唱 {%pho: [càng]} 歌什么的 // 因为马上  
{%pho: [sàng]} 离开校园了 // 这种机会也挺难得的 ///

En caso de que alguna pronunciación no exista en *pinyin*, también tratamos de marcarla con el signo fonético de *pinyin*.

### **6.2.2.3. Etiquetas para marcar la información prosódica**

**Pausas entonativas:** / sirve para indicar las unidades tonales. En términos generales, se pondrá una barra / cuando intuitivamente se perciba una pequeña pausa tonal, a pesar de que el hablante no haya terminado su unidad entonativa o enunciado.

\*YMJ: 第五种呢 / 叫白茶 /// 这个白茶呀 / 比较珍贵 //  
以前也称作 / 是给皇上啊 [/] 进贡给皇上的贡茶 ///

Como podemos observar, cuando el hablante llega a 呢 hace una pausa pero no ha terminado de hablar. Normalmente, la barra simple / se suele corresponder con una breve pausa con entonación ascendente, que indica que el hablante no ha finalizado su discurso (independientemente de si desde un punto de vista sintáctico la oración esté completa o no), mientras que con las dos barras // el hablante hace una pausa más marcada, a veces con entonación descendente, pero a continuación prosigue su unidad informativa.

**Final de enunciado:** estos cinco signos, ///, ?, !, ....., + (según sea asertivo, interrogativo, exclamativo, suspensión por conocimiento compartido o interrupción por parte de otro hablante) sirven para indicar el final del enunciado.

A diferencia del español, en chino sólo hay ? y !. Al principio de cada enunciado, no existe el uso de ¿ y ¡.

Los seis puntos (en chino, los puntos de suspensión son de seis puntos) indican el final de un enunciado con entonación suspendida. Es un enunciado que no se termina porque el hablante considera que su interlocutor ya ha inferido lo que quiere decir. Pongamos un ejemplo:

\*LXX: [<] <然后 / 对> /// 在吃方面 //  
尽量少在外面吃饭吧 // 虽然好像那些餐馆看起来都 .....

En este caso, el hablante considera que no es necesario terminar su enunciado porque se deduce lo que quiere decir. Así pues, los puntos suspensivos se ponen cuando la entonación suspendida guarda en sí misma una carga de significado que no se explicita pero que se da por supuesto. No hay que confundir los puntos

suspensivos con los alargamientos vocálicos: en este caso, el hablante no considera que su interlocutor pueda inferir el resto de la información, sino que el alargamiento, ya sea de un sonido vocálico o uno consonántico, en una estrategia de planificación del discurso, pero el hablante culminará su enunciado. Los alargamientos se señalan con el signo → a continuación del fonema alargado.

La triple barra (///) se indica para señalar el final de todos los demás tipos de enunciados, de la clase que sean, afirmativos, dubitativos, imperativos, etc.

• **Continuación de turno:** se pondrá este símbolo ⇢ a principio de turno para indicar que se continúa la unidad informativa del turno anterior, tras una interrupción o solapamiento.

\*LYM: 对 // 是在古巴很有名的 /// 就是 / 你只要随便拉一个古巴人 // 他都会给你跳一段 / SALSA /// 因为它的 +

\*LYH: 噢 / hhh {%act: assent} ///

\*LYM: ⇢节奏非常简单 // 我们在舞蹈课上也有学过 ///

#### 6.2.2.4 *Apoyos vocálicos*

Son signos de transcripción que se utilizan para expresar aquellas vocales que pronunciamos como por ejemplo “eh” y “ah”, con el fin de alargar el tiempo de planificación de habla. En nuestro corpus se van a transcribir de esta forma: &mm, &hao o &ne, etc.

\*GUY: &mm 我喜欢 [/] 我其实比较喜欢那种 // &mm 从西方的音乐改编成二胡曲的 // 那种乐曲 ///

En el caso de transcripciones de habla extranjera, se transcribirán los apoyos vocálicos propios de cada idioma. Por ejemplo, para el inglés se emplearán los siguientes: &uh, &um (variante nasalizada de la anterior), &er, &erm... Para el chino, &hao y &ne se usan en C-ORAL-CHINA.

\*JIJ: 没细看 // 不敢接近 hhh {%act: laugh} ///

\*ZHF: &nenenenenenene ///

\*YMJ: 现在在中国的 / 这个电视台上热播的 / 这个讲解这个论语的 // 这个 [/] 哎 / 就 [/] 就指的就是这本书 &hao ///

### 6.2.2.5 Interjecciones

Las interjecciones se transcriben con signos de exclamación !. No sirven para designar realidades sino únicamente para expresar el estado de ánimo del hablante.

\*SMW: [<] <对> hhh {%act: assent} /// 然后就是国内的酒 // 基本上就是如果我们喝的话 // 就是都喝 50 度以上的吗 /// <然后> +

\*DOY: [<] <啊> ! 你那么厉害 /// 我估计我喝一口就晕这儿了 ///

### 6.2.2.6 Los signos paralingüísticos

Los sonidos que no son lingüísticos pero que tienen significado, por ejemplo, de duda, de interrogación, etc. se reflejan como *hhh*. Algunos de estos signos paralingüísticos son:

laugh (risa)  
assent (asentimiento)  
click (chasqueo)  
cough (tos)  
negation (negación)  
doubt (duda)  
interrogation (interrogación)



blow (soplido)  
clearing of throat (carraspeo)  
yawn (bostezo)

A la hora de transcribir, en el texto de la transcripción se pondrá **hhh** mientras que el significado se pondrá en un comentario de tipo %act, tal y como ya hemos explicado más arriba.

\*WAL: &mm 我想西班牙人 [///] &mm 我想很 [/] 对于很多西班牙的同学来讲 // 那个 hhh {%act: clearing of the throat} 上这个课的目的呢 / 肯定想将来 / 有机会能够自己 / 比如在中国呀或者在西班牙也好 // 能够 hhh {%act: click} &mm 上一些新闻的看 [/] 上一些新闻网站啊 / 看一些实时的国内一些消息 ///

#### **6.2.2.7 Sonidos no reconocidos como signos lingüísticos**

• **Las tres equis:** se recurre a este signo de transcripción (xxx) cuando el transcriptor no entiende lo que ha dicho el hablante. Indica sonido ininteligible.

\*LYH: 哎 // 其实在这些 [///] 我想对于这些舞蹈形式 // 其中还是 [/] 可能是互相联系的 /// 是吧 ? 都不可能是单独存在 // 都可能 /// hhh {%act: click} / <XXX> +

\*LYM: [<] <那可不> ! 它就是 [///] 都是 [/] 舞蹈都是相通的吗 !

• **El ampersan:** es el signo (&) que tiene que preceder a palabras inconclusas o fragmentos.

\*WQZ: 有 [///] 一种是用手指涂和用 &海 [///] &mm 第二种呢 // 就是用海绵涂 ///

### 6.2.2.8 Reinicios y reformulaciones

Los signos para expresar reinicio o reformulación son dos: [/] y [///]. Se utilizan para aquellos casos de reprogramación del habla en los que el hablante se traba o se confunde al decir una palabra y necesita pronunciarla de nuevo. También sirve para aquellos casos en los que el hablante no se muestra satisfecho con lo dicho y prefiere comenzar otra vez. En el reinicio siempre hay dos partes:

\*SMW: 当然游了 // 不然能晒这么黑吗 ? 我们是 [/]  
我们是一大家子人一起去的 // 然后 / 所有人都下水了  
///

A la izquierda del signo de reinicio siempre se encuentra el fragmento desechado; mientras que a la derecha del reinicio se encuentra la opción que el hablante ha dado por válida. Dependiendo del grado de semejanza o variación que haya a un lado y a otro se elegirá entre [/] y [///]. Normalmente, se suele utilizar [/] para aquellos casos en los que a un lado y a otro existe o bien la misma cadena de habla, o bien dos palabras que a pesar de ser distintas tienen la misma categoría gramatical. Veamos un ejemplo de cada tipo:

\*SMW: 就是他们会调很好的装备 /// 然后我们就 [///]  
但是这个 [/] 这个地方你一个人是打不了的 // 你就是大  
大家要组队打 ///

Como vemos, en estos ejemplos marcados con [/] hay repetición pero no se produce una ruptura de la sintaxis de la oración. Mientras que cuando se utiliza el signo con tres barras [///], la reprogramación del habla es completa.

### 6.2.2.9 Interrupciones

Existen dos tipos de interrupciones.

• **La autointerrupción (=):** esto es, cuando el hablante corta su discurso de una forma brusca para hablar de otro tema completamente distinto al que estaba tratando anteriormente. La diferencia entre = y [//] es temática, es decir, que mientras = implica un cambio de tema dentro del discurso, [//] no.

\*SMW: [<] <没有我们 [/] 我们>那个 = 他们好像就是说 / 去 [/] 去 FIESTA 之前 // 他们就是要先在家里先喝一下 /// 他们说那个叫什么 BOTELLON ///

• **La interrupción** de un participante a otro cuando el primero no ha concluido su enunciado. Esto se marca con el signo + en el punto de la interrupción, para indicar que el enunciado está inconcluso.

\*LYH: 古巴主要的支柱产业 +

\*LYM: <你可以猜一下>

#### **6.2.2.10 Pausa**

El signo # se utiliza para marcar una pausa prolongada. Es un silencio, un espacio temporalmente muy perceptible y que no se pueda considerar una pausa prosódica.

\*CRJ: 如果一些不是很常用的 // 或者是书面语 [/] 书面语色彩特别强的词语 // 可以暂时不用 [/] 不用掌握 #

#### **6.2.2.11 Solapamiento**

El solapamiento se produce cuando dos hablantes hablan a la vez. El texto solapado aparece entre los signos <>. En el siguiente turno, [<] indica que se solapa con el anterior.

\*LYH: <&mm >

\*LYM: [<] <他> [/] 他们都有自己的想法 ///

### 6.3 Alineamiento

El alineamiento consiste en la sincronización del texto transcrito con el sonido original, de forma segmentada, es decir, o bien por turnos conversacionales o bien por enunciados. C-ORAL-CHINA está alineado por enunciados, siguiendo la metodología del LLI-UAM.

Tanto las transcripciones como el alineamiento se llevan a cabo con Transana (versión 2.12), un programa de transcripción y análisis de datos creado en la Universidad de Wisconsin y disponible de forma gratuita a través de su página web<sup>10</sup>.

En nuestra opinión, este programa presenta tres ventajas básicas:

- **Simultaneidad de operaciones:** Transana se presenta en un formato que permite mantener activas simultáneamente varias ventanas, entre las que se incluyen la ventana de texto y la de sonido (también está preparado para video, en caso de que la grabación incluya imágenes). De esta forma, se puede operar en la ventana de texto al mismo tiempo que se escucha el sonido. El programa se compone de cuatro ventanas: arriba, en la parte izquierda, aparece la ventana de sonido con la imagen de la onda sonora e información sobre el tiempo total de duración de la grabación, o el tiempo actual de la transcripción; debajo de esta ventana encontramos la de texto, con una pequeña barra de herramientas para la edición; en la parte superior derecha tenemos la ventana de video; y debajo de ella está la ventana de datos, donde almacenamos los distintos archivos de sonido y texto.

---

<sup>10</sup> [en línea]: <[www.transana.org](http://www.transana.org)> [consulta: 25 de noviembre de 2010]

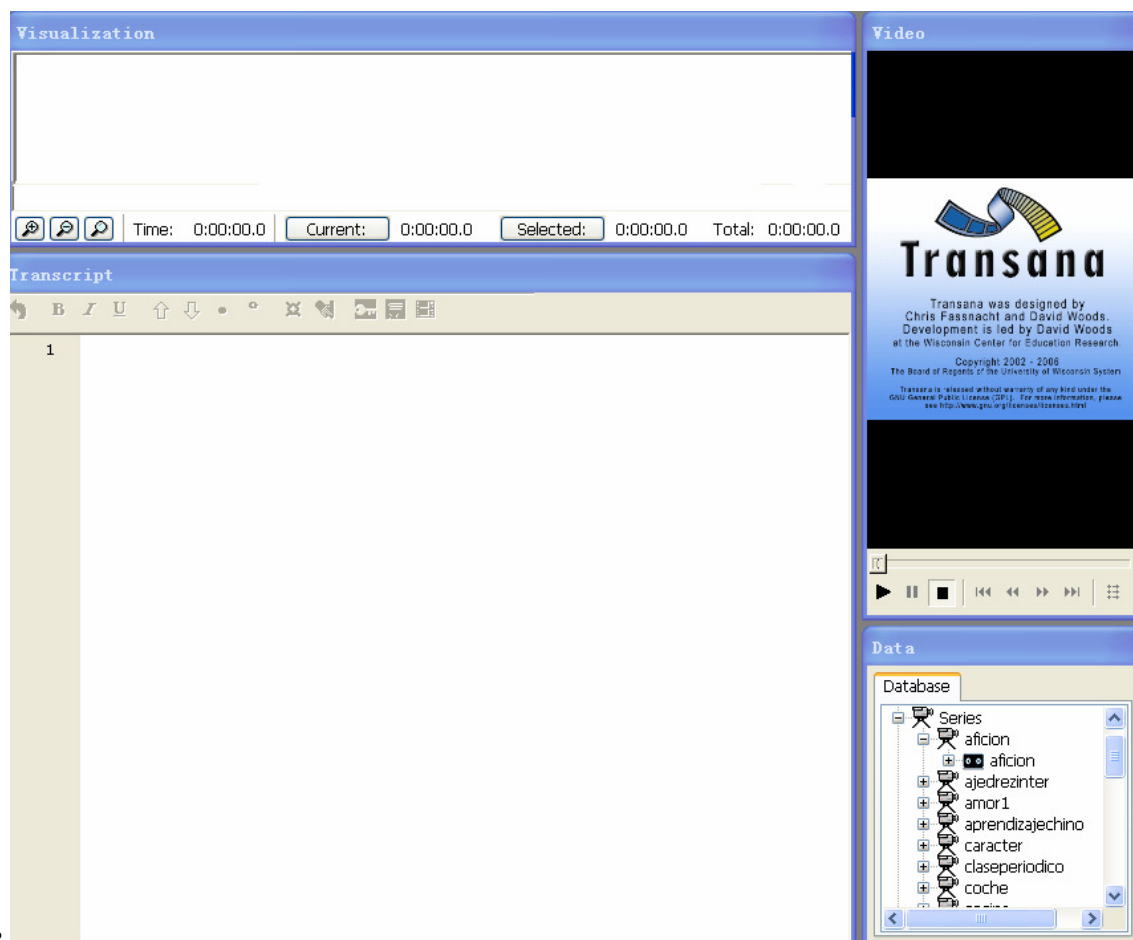


Figura 6.3 Programa de transcripción Transana

• **Ergonomía:** Transana dispone de un número de atajos en el teclado con los que se facilita la labor del transcriptor:

Ctrl+S	Parada/marcha; retrocede automáticamente dos segundos
Ctrl+D	Parada/marcha; sin retroceso
Ctrl+A	Reinicio diez segundos atrás
Ctrl+F	Reinicio diez segundos adelante
Ctrl+P	Comienza en el principio de la grabación

• **Alineamiento simultáneo:** Transana permite alinear el sonido durante el mismo proceso de transcripción. Utilizando el atajo Ctrl+D se para el sonido en el momento preciso de la finalización de cada enunciado, y se marca con el símbolo ■ (Ctrl+T) destinado a esta

tarea, sincronizando o alineando el segmento de sonido con su correspondiente fragmento textual.

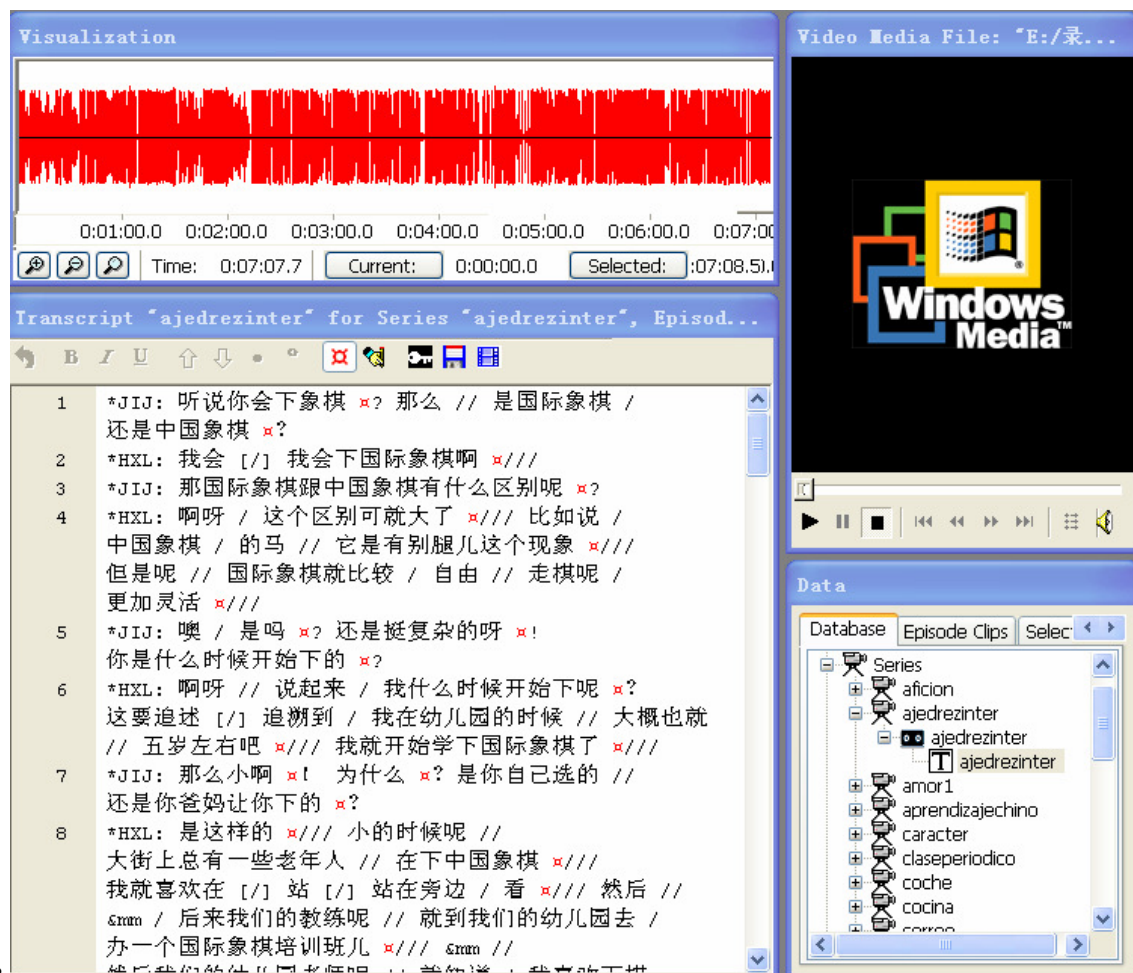


Figura 6.4 Alineamiento en Transana

El texto que produce Transana se copia y pega en un documento Word. Este documento hay que guardarlo con extensión .txt (guardar como texto sin formato). Sobre este documento .txt se aplicarán tanto el programa de conversión al formato XML como el transcriptor de *pinyin*.

## 6.4 Transcripción de pinyin

Entre las grabaciones de C-ORAL-CHINA, hemos seleccionado unas cuantas para hacer la transcripción de *pinyin*. El transcriptor de *pinyin* se aplica al texto de forma automática, posibilitando la transcripción del mismo en poco tiempo.

Una vez que tenemos el texto transcrito y alineado (con las marcas de tiempo), se copia y se guarda como archivo de texto, txt. Sobre este archivo de texto aplicaremos el programa de transcripción de *pinyin*. Hay unas páginas web exclusivas para hacer este trabajo. También hay *software* exclusivo, por ejemplo, el que hemos usado en C-ORAL-CHINA, KTestpinyin.exe.

<http://pinyin.911cha.com/>

<http://xh.5156edu.com/conversion.html>

<http://py.kdd.cc/>

<http://qq.ip138.com/wb/wb.asp>

[http://tool.chinaz.com/Tools/Word\\_Spell.aspx](http://tool.chinaz.com/Tools/Word_Spell.aspx)

<http://hanyu.iciba.com/pinyin>

[http://tool.webmasterhome.cn/Word\\_Spell.asp](http://tool.webmasterhome.cn/Word_Spell.asp)

Con posterioridad, es muy necesaria una revisión manual para asegurar la fiabilidad y corregir posibles errores. Por ejemplo, en *pinyin*, existen tono ligero y *duoyinzi* (多音字: son aquellos que se pueden pronunciar de manera diferente). Veamos un ejemplo: el carácter “乐” tiene diferentes pronunciaciones. “乐趣” se pronuncia como “lèqù”, mientras “音乐”, como “yīnyuè”. En muchos casos, los programas no pueden distinguirlos bien, hay que revisarlos y corregirlos de manera muy detallada.

## 6.5 Conversión XML

El texto de salida en Transana se puede exportar a cualquier otro editor de texto, como por ejemplo Microsoft Word o Edit Plus,

simplemente seleccionando el texto, copiando y pegándolo en el documento elegido. En nuestro caso, nos interesaba convertir el formato de texto de Transana en un documento en XML. Sin embargo, las marcas de tiempo establecidas en Transana para alinear se convierten en cifras numéricas al exportar el texto a otro tipo de documento. Esas cifras numéricas expresan el tiempo exacto del alineamiento de texto y sonido en milisegundos, quedando de la siguiente manera como la figura 6.5:

\*AZY: 听你说了这 {%pho: [zè]} 么多 / 我感觉你是一个性格很开朗的人 ///Q<4957>

\*YUF: 还好吧我 / 我觉得 ///Q<6439>

\*AZY: 我很喜欢你这种性格的人 # 无论什么时候见到你 // 总是那么神采奕奕的 ///Q<13999>

\*YUF: 可能是我没有太多的烦心事吧 ///Q<16482> 你看 / 有的时候 / 发生些什么事情 / 我都会以一种自己的方式来解决 ///Q<21732> 然后呢 / 想通 [/] 想通了也就好了 ///Q<24897> 其实 / 性格这种问 [/] 这种事 [/] 事情吧 / 不是说你觉得它好就是好的 ///Q<31006> 一定要是有个自己的磨合 / 找到一个适合自己发泄的方式 ///Q<35495> 然后 [///] 其实我觉得 / 应该把最开心的一面留给大家 / &mm / 让大家觉得 / 自己 [///] &mm 让大家觉得身边都是充满了开心与 [/] 与欢乐的 ///Q<46052> 如果每天愁眉苦脸的话 / 可能身边也没有那么多的朋友吧 .....Q<50062>

\*AZY: hhh {%act: assent} &mm 看来是的 ///Q<52169>

Figura 6.5 Salida del alineamiento en transana

Es necesario convertir el formato de salida de Transana al formato convencional de C-ORAL-ROM, es decir, desglosando el texto en enunciados limitados por una marca de tiempo al inicio y otra al final



de los mismos. Así pues, se hizo un programa que realizara el cambio de forma automática para que el documento final mostrase el mismo aspecto y formato que las transcripciones alineadas de C-ORAL-ROM, quedando así:

```
<UNIT speaker="AZY" startTime="0" endTime="4.957">听你说了
这 { %pho: [zè] } 么多 / 我感觉你是一个性格很开朗的人
///</UNIT>
```

```
<UNIT speaker="YUF" startTime="0" endTime="6.439">还好吧我
/ 我觉得 ///</UNIT>
```

```
<UNIT speaker="AZY" startTime="6.439" endTime="13.999">我很
喜欢你这种性格的人 # 无论什么时候见到你 // 总是那么神采奕奕
的 ///</UNIT>
```

```
<UNIT speaker="YUF" startTime="13.999" endTime="16.482">可
能是我没有太多的烦心事吧 ///</UNIT>
```

```
<UNIT speaker="YUF" startTime="16.482" endTime="21.732"> 你
看 / 有的时候 / 发生些什么事情 / 我都会以一种自己的方式来解决
///</UNIT>
```

```
<UNIT speaker="YUF" startTime="21.732" endTime="24.897"> 然
后呢 / 想通 [/] 想通了也就好了 ///</UNIT>
```

```
<UNIT speaker="YUF" startTime="24.897" endTime="31.006"> 其
实 / 性格这种问 [/] 这种事 [/] 事情吧 / 不是说你觉得它好就是
好的 ///</UNIT>
```

```
<UNIT speaker="YUF" startTime="31.006" endTime="35.495"> 一
定要是有个自己的磨合 / 找到一个适合自己发泄的方式 ///</UNIT>
```

```
<UNIT speaker="YUF" startTime="35.495" endTime="46.052"> 然
后 [///] 其实我觉得 / 应该把最开心的一面留给大家 / &mm /
让大家觉得 / 自己 [///] &mm 让大家觉得身边都是充满了开心
```

```

与 [/] 与欢乐的 ///</UNIT>

<UNIT speaker="YUF" startTime="46.052" endTime="50.062"> 如
果每天愁眉苦脸的话 / 可能身边也没有那么多的朋友吧
.....</UNIT>

<UNIT speaker="AZY" startTime="50.062" endTime="52.169">hhh
{%act: assent} &mm 看来是的 ///</UNIT>

```

Figura 6.6 Ejemplo de salida a formato XML

La etiqueta <UNIT> incluye un solo enunciado, especificando el tiempo de inicio y el de finalización en formato XML. En Transana, la primera marca de tiempo no puede corresponderse con 0, ya que ha tenido que pasar al menos un segundo para ponerla. Aquí, la primera marca en el ejemplo es <4957>, que indica el final del primer enunciado en milisegundos. El programa de cambio a formato C-ORAL-ROM establece el inicio del primer enunciado en 0 y el final en “4957”; esta cifra será a su vez el tiempo de inicio del siguiente enunciado, cuyo tiempo de finalización será de nuevo el tiempo de inicio de la siguiente y así sucesivamente.

## 6.6 Resumen

Todos los símbolos se escribirán dejando un espacio en blanco a ambos lados del mismo. En la tabla 6.7 se resumen las convenciones usadas en C-ORAL-CHINA.

Símbolo	Descripción	Etiqueta XML	Ejemplo
*ABC:	Iniciales del participante	<Turn> <Name> xxx </Name> <Says> zzz </Says> </Turn>	*SMW: &mm 我去 海边玩了 ///
/	Interrupción prosódica terminal y no autónoma	<Tone_Unit Type="standard" />	*DOY: 啊呀 / 亲爱的 // 你怎么晒黑了 ?

//	Interrupción prosódica terminal autónoma	no y <Tone_Unit Type="complex" />	*HXL: &mm / 比较好玩儿 // 还更加生动 ///
///	Interrupción prosódica terminal; entonación descendente	<Utterance Type="enunciation">	*HXL: 小的时候呢 // 大街上总有一些老年人 // 在下中国象棋 ///
?	Interrupción prosódica terminal; entonación interrogativa	<Utterance Type="interrogation"> xxx </Utterance>	*JIJ: 听说你会下象棋 ?
!	Interrupción prosódica terminal; entonación exclamativa	<Utterance Type="exclamation"> xxx </Utterance>	*JIJ: 天哪 // 你好厉害呀 !
.....	Interrupción prosódica terminal; entonación suspendida	<Utterance Type="suspension">	*LXX: [<] <然后 / 对> /// 在吃方面 // 尽量少在外面吃饭吧 // 虽然好像那些餐馆看起来都 .....
=	Autointerrupción	<Utterance Type="self-interruption">	*GLT: 然后因为 = 下班就不行了 //
¬	Continuación de turno tras interrupción o solapamiento	<Continues />	*LYM: <因为>+ *LYH: [<] <这要> *LYM: ¬还是在卡斯特罗这位老同志的领导之下 ///
+	Interrupción	<Utterance Type="interruption">	*LYM: 或者只是 [/] 他只是想打一个招呼 + *LYH: <对>
[/]	Reformulación o reinicio léxico o sintagmático	<Tone_Unit Type="partial_restart" />	*LYH: 那么大的 [/] 那么大的芒果 !
[[/]]	Reformulación sintáctica	<Tone_Unit Type="total_restart" />	*LYH: [<] <噢 // 据说古巴> 在水果还有果汁这方面 / 特别 [[/]] 就是有限制是吧 ?

< > [<]	Solapamiento	<Overlap> xxx </Overlap>	*LYH: <是吧> *LYM: [<] <你走> 在街上 //
&	Palabra incompleta	<Fragment> xxx </Fragment>	*WQZ: 有 [///] 一种 是用手指涂和用 & 海 [///] &mm 第二种 呢 // 就是用海绵涂 ///
hhh	Signos paralingüísticos	<Non_Linguistic />	*LYH: 可以说 // 也 # hhh {%act: click} 是一种引领着 / 人 们的一种思想的 / 一个 [///] 这样一个 人物 ///
xxx	Fragmentos ininteligibles	<Unintelligible />	*LYH: 因为他已经 / 几乎成了 / 现在 XXX 一种精神的象 征 ///
&hao/&ne	Apoyos vocálicos	<DM> &eh </DM>	*YMJ: 这个 [/] 哎 / 就 [/] 就指的就是这 本书 &hao ///
#	Pausa no prosódica	<Pause />	*XUN: &mm 这个是 不一样的 /// &mm 北 # 京的标准
%alt:	Errores fenómenos normativos o no	<Notes type="xxx"> zzz </Notes>	*LIC: 啊 / 说到排 球 / 我也蛮感兴趣 的 /// 您们 {%alt: el dato correcto es 您几 位} 一般都什么时 间去打呀 ?
%com:	Comentarios	<Notes type="xxx"> zzz </Notes>	*CRJ: hhh {%com: open the book} 到第 二十课 ///
%act:	Paralenguaje	<Notes type="xxx"> zzz </Notes>	*CRJ: hhh {%act: clearing of throat} 上 课的时候 //

Tabla 6.7 Convenciones de la transcripción

## **7. Aplicación a la enseñanza de lengua segunda mediante el desarrollo de recursos basados en el corpus: propuestas didácticas**

Como hemos comentado en el capítulo 3, los corpus constituyen una herramienta útil, renovadora y práctica en la enseñanza de lengua segunda. Las aplicaciones concretas nos permiten realizar actividades de diverso tipo encaminadas al desarrollo de la conciencia lingüística y a descubrir cómo funcionan diferentes elementos. En este capítulo, se plantean propuestas didácticas sobre la enseñanza de aspectos léxicos, gramaticales, discursivos y fonemas.

Seguimos usando la herramienta de concordancias desarrolladas por el LLI-UAM. Gracias a ella, podemos formular una búsqueda, recuperar rápidamente ejemplos con contexto y escuchar las grabaciones.

Teniendo en cuenta diferentes necesidades de concordancias, en el corpus C-ORAL-CHINA se adoptan dos métodos para hacer concordancias de muestras. Un método es el de buscar el contexto y la grabación entera. Cuesta más tiempo de trabajo, pero permite la comprensión total del texto y localizar construcciones no consideradas. El otro método es el de hacer aparecer todo el enunciado por medio de introducir alguna palabra o locución de éste, gracias a que todas las muestras han sido cortadas ya en enunciados y sus textos han sido alineados con el sonido correspondiente.

Dadas las características de lengua oral de C-ORAL-CHINA y el nivel de dificultad de aquellos documentos de los que hemos seleccionado, las muestras oscilan entre un nivel intermedio y avanzado.

## **7.1 El uso de corpus en la enseñanza del léxico**

En comparación con corpus generales, una de las ventajas de nuestro corpus es que el vocabulario se concentra en unos temas fijos de la vida cotidiana. En la transcripción, ya hemos filtrado el contenido y quitado la parte inconveniente para la enseñanza.

Para consultarlo, los interesados pueden visitar la página web:

<http://drusila.llf.uam.es/lab/?q=c-oral-china>.

Usuario: visitante

Contraseña:

En ésta, la herramienta de concordancias es muy fácil de manejar, incluso para la gente que no tenga muchos conocimientos informáticos. Por medio de una búsqueda rápida y sencilla, los alumnos pueden familiarizarse con el vocabulario según su necesidad y conocer el uso adecuado de las palabras o locuciones en diferentes contextos, incluso el uso más nuevo que algunos diccionarios no contienen todavía.

Por ejemplo, si queremos aprender las palabras relacionadas con el catarro, en C-ORAL-CHINA, podemos encontrar un diálogo entre paciente y médica.

- 1 0-4.352 MÉDICA: 下一号 ///
- 2 4.352-6.478 PACIENTE: 哎 / 大夫您好 ///
- 3 6.478-8.091 MÉDICA: 噢 / 来看病 &hao // 请坐 ///
- 4 8.091-8.891 PACIENTE: &mm 谢谢 ///
- 5 8.891-9.826 MÉDICA: 多大年龄了 ?
- 6 9.826-11.242 PACIENTE: 24 ///
- 7 11.242-13.032 MÉDICA: 噢 / 哪里不舒服啊 ?
- 8 13.032-19.934 PACIENTE: &mm / 我这两天有点儿流鼻涕 / 然后还打喷嚏 / 另外嗓子也有点儿疼 ///
- 9 19.934-20.866 MÉDICA: 发烧吗 ?

10 20.866-25.232 PACIENTE: &mm / 刚才在外面护士给量的 / 是 38 度多 ///

11 25.232-26.123 MÉDICA: 噢 ///

12 26.123-32.096 MÉDICA: 根据你的症状啊 / 还有你的体温呢 / &mm 估计你这是得了上呼吸道感染 ///

13 32.096-33.796 MÉDICA: 啊 ///

14 33.796-35.853 MÉDICA: 那个还有别的不舒服吗 ?

15 35.853-39.256 PACIENTE: &mm 吃饭也不是太有胃口 ///

16 39.256-39.675 MÉDICA: 噢 ///

17 39.675-41.265 MÉDICA: 来我看看你的嗓子 ///

18 41.265-42.5 PACIENTE: &ah hhh {%act: open the mouth} ///

19 42.5-45.703 MÉDICA: 噢 // 好 / 嗓子有点儿红啊 ///

20 45.703-46.799 MÉDICA: 扁桃腺不大 ///

21 46.799-53.237 MÉDICA: 那个 / 去做一个化验吧 // 看看 &mm 血项怎么样 // 好不好 ?

22 53.237-54.07 PACIENTE: &mm 好 hhh {%act: assent} ///

23 54.07-54.952 MÉDICA: 我给你开一个单子 ///

24 54.952-55.494 MÉDICA: <啊>

25 54.952-55.494 PACIENTE: [<] < &mm > hhh {%act: assent} ///

26 55.494-58.954 MÉDICA: &mm 给你 / 拿这单子上化验室做化验 ///

27 58.954-61.436 MÉDICA: &mm 做完化验 / 再拿这单子来过来找我 hhh {%com: give the paper to the patient} ///

28 61.436-62.632 PACIENTE: &mm hhh {%act: assent} 好 hhh {%com: take it} ///

29 62.632-67.612 PACIENTE: &mm 大夫这个是那个化验室的单子 hhh {%com: give the paper to the doctor} ///

30 67.612-68.643 MÉDICA: 噢 / 我看看噢 ///

31 68.643-75.763 MÉDICA: &mm hhh {%com: take the paper} 从这个单子上看吧 / &mm 血项不太高 // 那个还是病毒感染 ///

32 75.763-76.356 PACIENTE: &mm hhh {%act: assent} ///

33 76.356-79.45 MÉDICA: 所以 / 我 [/] 那个我给你开点药 ///

34 79.45-80.201 PACIENTE: < &mm hhh {%act: assent}>

35 79.45-80.201 MÉDICA: [<] <好吧> ?

36 80.201-84.336 MÉDICA: &mm 开的药主要是抗感冒的 ///

37 84.336-85.011 MÉDICA: <噢>

38 84.336-85.011 PACIENTE: [<] < &mm hhh {%act: assent}> ///

39 85.011-86.693 MÉDICA: &mm 中药和西药都有 ///

40 86.693-92.396 MÉDICA: 这个中药呢 / 是咱们开的感冒冲剂 / 是要冲着喝的 // 一天两次 / 一次两袋儿 ///

41 92.396-93.162 MÉDICA: <&hao>

42 92.396-93.162 PACIENTE: [<] < &mm hhh {%act: assent}> ///

43 93.162-97.023 MÉDICA: 西药呢 / 是给你拿的是 / &mm 速效感冒胶囊 ///

44 97.023-97.912 PACIENTE: &mm hhh {%act: assent} ///

45 97.912-99.913 MÉDICA: 一天三次 / 一次一粒 ///

46 99.913-101.428 PACIENTE: &mm hhh {%act: assent} ///

47 101.428-104.052 MÉDICA: 消炎药/ 我就不给你开了 / 噢 ///

48 104.052-108.499 MÉDICA: 因为你这个血项不太高 / 是病毒感染的 / 咱们不用吃消炎药了 ///

49 108.499-110.561 MÉDICA: 你大量地喝热水 / 也可以 &hao ///

50 110.561-113.888 MÉDICA: 再有 / 我给你开点退烧药 / 这是 ABC 的 ///

51 113.888-114.497 PACIENTE: &mm hhh {%act: assent} ///

52 114.497-119.28 MÉDICA: 吃了药以后 / 多 [/] 大量地喝热水 / 烧就能退了 &hao ///

53 119.28-122.444 PACIENTE: 这些药是饭前吃还是饭后吃啊 ?

54 122.444-125.156 MÉDICA: 这个不一定 // 没关系 / 饭前饭后都可以 ///

55 125.156-126.903 PACIENTE: 噢 hhh {%act: assent} // 要分开吗 ?

56 126.903-128.554 MÉDICA: 中药和西药要分开

57 128.554-129.654 MÉDICA: // 二十分钟 // 间隔二十分钟 ///

58 129.654-130.262 MÉDICA: <吃>

59 129.654-130.262 PACIENTE: [<] < &mm hhh {%act: assent}> ///

60 130.262-134.207 MÉDICA: &mm &mm 吃饭要注意什么 ?

61 134.207-139.778 MÉDICA: 注意点儿那个不要吃辛辣的东西 &hao // 刺激的 / 酸辣的尽量避免一下 ///

62 139.778-140.596 PACIENTE: 都不能吃 ///

63 140.596-141.767 PACIENTE: <是吧 &mm >

64 140.596-141.767 MÉDICA: [<]<对> ///



65 141.767-148.907 MÉDICA: 行 / 给你拿这个方子 // 先去一楼大厅去划价室划价 ///

66 148.907-149.291 PACIENTE: &mm hhh {%act: assent} ///

67 149.291-151.165 MÉDICA: 完了 / 再上药房拿药就可以了 ///

68 151.165-151.885 PACIENTE: &mm hhh {%act: assent} ///

69 151.885-152.29 MÉDICA: 好吧 ?

70 152.29-152.804 PACIENTE: hhh {%act: assent} 好 ///

71 152.804-157.36 MÉDICA: &mm 过几天 / 如果不见好转 // 那个记得来复查一下来 ///

72 157.36-160.086 PACIENTE: &mm 好 hhh {%act: assent} ///

73 160.086-169.173 PACIENTE: 另外我还想问一下 / 就是我觉得自己经常感冒 // 这种情况是不是应该接种一下那种流感疫苗 ?

74 169.173-170.604 MÉDICA: 嗯 / 是应该接种 ///

75 170.604-174.328 MÉDICA: 这个接种的时间 / 应该是在每年的 10 月份接种 ///

76 174.328-184 MÉDICA: 这个流感疫苗是非常好的一个疫苗儿 // 能 / 预防自己不得感冒 / 而且也 [/] 也能对他人进行保护 ///

77 184-187.266 PACIENTE: 每年只有 10 月份才能接种啊 ?

78 187.266-196.426 MÉDICA: 对 hhh {%act: assent} / 10 月份接种以后 / 半个月到 20 天产生抗体 / 能预防冬天的和 / 春季的流感 / 流行 ///

79 196.426-200.843 PACIENTE: 噢 hhh {%act: assent} / 行 / 那我就等明年 10 月份过来 ///

80 200.843-201.748 MÉDICA: &mm 好 hhh {%act: assent} ///

81 201.748-202.498 PACIENTE: &mm 好 hhh {%act: assent} ///

82 202.498-203.403 PACIENTE: 谢谢大夫 ///

83 203.403-204.154 MÉDICA: 好 / 再见 !

84 204.154-204.921 PACIENTE: &mm 再见 !

Figura 7.1 Ejemplo de C-ORAL-CHINA

En este documento, podemos sacar un listado de léxico médico para que los alumnos practiquen.

Síntomas: 流鼻涕 (caérsele a alguien los mocos), 打喷嚏 (estornudar), 嗓子疼 (doler a alguien la garganta), 发烧 (tener la

fiebre), 上呼吸道感染 (infección del tracto respiratorio superior), 病毒感染 (infección viral), 流感 (la gripe), etc.

Tratamientos: 量体温 (tomar la temperatura), 血项化验 (análisis de sangre), 开药 (recetar), 拿药 (ir por medicina), 吃药 (tomar pastillas), 复查 (revisión), 接种流感疫苗 (vacunación contra la gripe), etc.

Medicinas: 中药 (medicina tradicional de China), 西药 (medicina occidental), 消炎药 (medicina anti-inflamatorios), 退烧药 (antipirético), etc.

Son expresiones y estructuras muy frecuentes usadas por los nativos. No aparecen solas, sino en un contexto concreto. Los alumnos pueden deducir el significado de algunas palabras desconocidas por el contexto y su conocimiento del mundo.

También podemos usar la herramienta para hacer una comparación semántica detallada entre diferentes palabras.

Entramos en la página web. Sale esta pantalla:



Figura 7.2 Pantalla principal de la página web

Seleccionamos *Buscador*, sale la siguiente página:

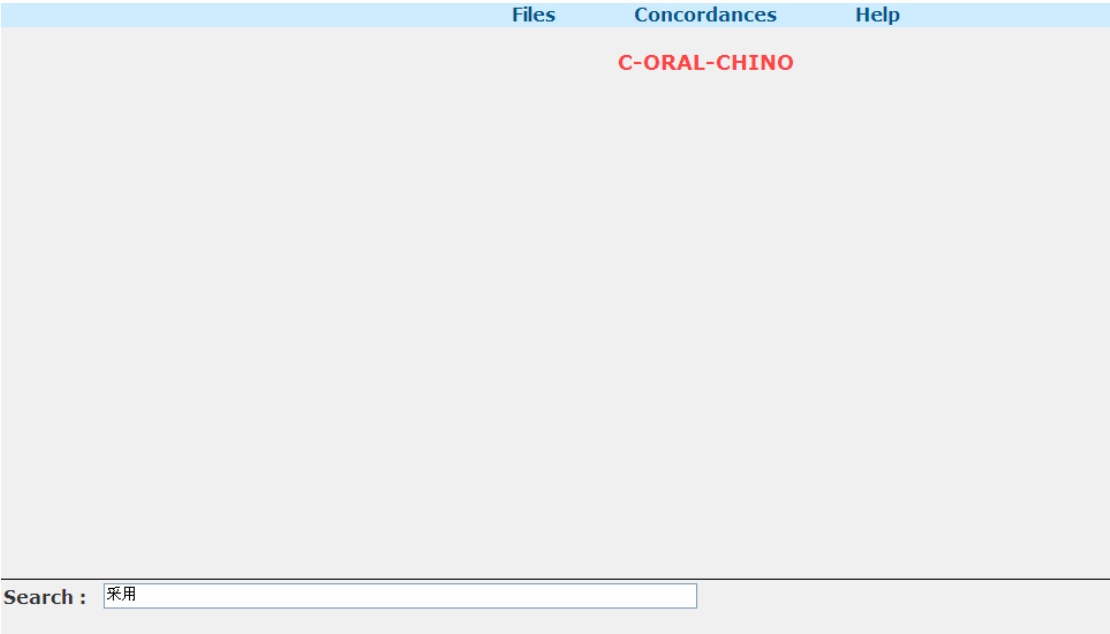


Figura 7.3 Buscador de C-ORAL-CHINA

Rellenamos en el blanco el contenido que queremos buscar. En este caso, son dos palabras “感动” y “激动” .



Figura 7.4 Ejemplos de la palabra buscada “感动”



Figura 7.5 Ejemplos de la palabra buscada “激动”

Lo que obtenemos es una pantalla con enunciados que contienen la palabra buscada. En las figuras 7.4 y 7.5, al lado de cada enunciado, aparece un botón que al accionarse nos reproduce el fragmento seleccionado. Es decir, podemos leer la transcripción a la vez que escuchamos el sonido. A la derecha, se ofrecen las informaciones sociolingüísticas del hablante y número del enunciado en el documento original.

Si los alumnos no entienden algún enunciado o tienen mucho interés, pueden consultar directamente ese documento completo para poder entender mejor. Por ejemplo, el primer enunciado de “感动” viene de chin02. Volvemos a la página principal. Seleccionamos *Transcripciones*. Nos sale la siguiente pantalla como la figura 7.6.

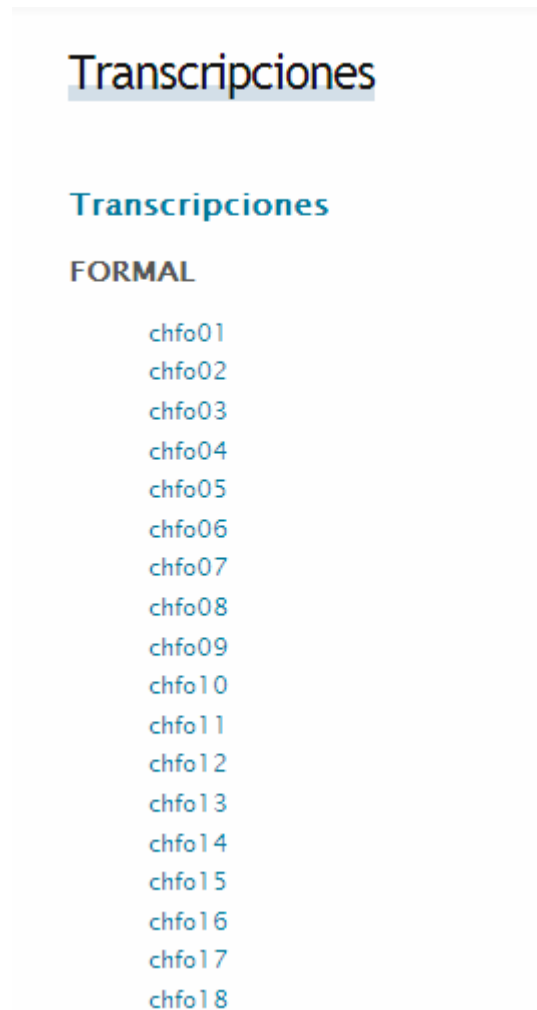


Figura 7.6 Ficheros de *Transcripciones*

Seleccionamos el fichero original del primer enunciado, y entramos en ese fichero chin02.

114 348.67-351.36 JIJ: 你有没有看过那个日本漫画 ?  
 115 351.36-352.512 JIJ: 有一个棋魂 ///  
 116 352.512-357.226 JIJ: 哎 / 不对 // 那个不是讲下棋的事 [ ] 下象棋的事情 // 是在讲下围棋的事情 ///  
 117 357.226-361.474 JIJ: 就是里面那个主人公 // 他把下围棋当成自己的生命一样 ///  
 118 361.474-367.184 JIJ: 然后在死了之后 // 他的魂魄都寄托在那个围棋之上 // 让人看了很感动 ///  
 119 367.184-370.013 HXL: 啊呀 // 对不起啊 // 我不喜欢看日本漫画 hhh {%act: laugh} ///  
 120 370.013-372.169 JIJ: 啊呀 // 那你有机会看一下吧 ///  
 121 372.169-374.469 JIJ: 因为棋 // 我觉得 // 都是相通的吧 ///

Figura 7.7 Contenido del fichero chin02

Podemos endenterlo en el contexto.

Si no estamos decididos y queremos saber de qué trata cada fichero, podemos entrar en *Buscador* y seleccionar *Files*. En seguida, tenemos informaciones resumidas de cada fichero como la Figura 7.8 .

Files		Concordances	Help	
FILES				
chfo01	chfo02	chfo03	chfo04	chfo05
Textbooks, prep, review, customs, grammar FORMAL 22' 14" 4348 caracteres	newspaper, spanish students, learning chinese FORMAL 15' 25" 3663 caracteres	Confucius, teaching, the Analects of Confucius  FORMAL 10' 11" 2096 caracteres	Flood control FORMAL 06' 08" 1358 caracteres	activity of Fu Xishi FORMAL 02' 28" 530 caracteres

Figura 7.8 Informaciones resumidas de todos los ficheros de C-ORAL-CHINA

Si vemos que hay algún tema que nos interesa, podemos pinchar el título para escucharlo o leerlo.

Al final, con tantos ejemplos vivos, podemos concluir que “感动” significa que algo impresiona a alguien y se emociona. “激动” quiere

expresar el estado de excitarse, conmoverse o agitarse.

Otro ejemplo es un par de sinónimos “采用” y “采纳”. Escribimos “采用”.

Pasamos a la página siguiente:

Files	Concordances	Help
采用		
采用芦柴灰 / 堵住了冒水的裂缝 //		YMJ, profesor, woman, D years, chfo09,72
&mm 但是呢 / 我们就 / &mm 采用一站式的这个购物环境 //		XUN, staff, man, A years, chin17,33
奥运会足球比赛场地 / 采用天然草坪 //		XYZ, student, man, A years, chme06,27
长105米 //		
宽68米 //		
对 //		
然后 / 像我同学他们都会采用网购的方式 //		XYZ, , , years, chme20,32
&mm 逛街走路蛮累的 hhh {%act: laugh} //		
若您财付通帐户中余额不足支付 //		SYC, student, woman, A years, chme20,43
推荐您采用财付通一点通关联支付 //		
Search : <input type="text"/>		

Figura 7.9 Ejemplos de la palabra buscada “采用”

De la misma manera, podemos sacar los ejemplos de la otra palabra “采纳”.



Figura 7.10 Ejemplos de la palabra buscada “采纳”

Después de leer detalladamente los ejemplos, podemos hacer la comparación. Notamos la diferencia entre las dos palabras. La primera se refiere a utilizar o emplear algo, y la segunda significa adoptar o aceptar alguna propuesta.

## ***7.2 El uso de corpus en la enseñanza de la gramática***

Igual como en la enseñanza del léxico, el corpus también puede usarse para la enseñanza de la gramática. Tradicionalmente, el estudio de la gramática se basa en la lengua escrita. La compilación de corpus de habla espontánea nos trae un recurso muy valioso para conocer y estudiar las características de la gramática de la lengua hablada.

El profesor puede diseñar un programa de gramática y extraer muestras. A continuación, veremos cómo usar los ejemplos del corpus para explicar algunos puntos gramaticales característicos del chino.



Primero, va la explicación del punto gramatical en que se subraya la diferencia entre el chino y el español. En comparación con otros tipos de gramáticas, proporcionamos el contexto de cada oración para facilitar la comprensión de los alumnos. Y luego viene la transcripción de *pinyin*. Hace falta explicar que teniendo en cuenta que muchos alumnos no saben el principio de la segmentación de palabras de chino, aquí el *pinyin* se presenta separado por carácter, en lugar de por palabra. A continuación, va la traducción al español. Los alumnos pueden notar la similitud y diferencia entre los dos idiomas, y comprobar si su comprensión es correcta o no. Por último, se indica el nombre del documento para que los interesados puedan consultar todo el texto.

### **Clasificador**

Es una categoría de palabras que indica la cantidad o categoría de personas, objetos y acciones. Se sitúa detrás de los numerales y delante de los sustantivos. Su uso depende del sustantivo que le sigue. En muchos casos cuando se traduce del chino al español no aparece. Otras veces se corresponde con el artículo interminado *uno/una* o los numerales.

### **La función del clasificador:**

1. El uso de distintos clasificadores puede expresar las características morfológicas de diferentes aspectos del mismo sustantivo. Por ejemplo, cuando decimos “一张地图”, el clasificador “张” indica que el mapa es único, mientras que si decimos “一本地图”, “本” nos indica el conjunto de mapas.
2. Puede enfatizar las características morfológicas<sup>11</sup> principales de la entidad descrita. Por ejemplo, cuando usamos “粒” o “颗” como clasificador de la legumbre, enfatiza la forma peculiar de los granos de la legumbre. Muchos clasificadores han captado la parte más

---

<sup>11</sup> En el sentido de apariencia física, no en el sentido lingüístico de propiedades gramaticales como el tiempo, número o persona.

llamativa y representativa de la morfología exterior de los objetos estáticos y dinámicos.

3. Igualmente, el uso del clasificador puede distinguir los objetos que pertenecen al mismo tipo pero tienen diferentes morfologías exteriores. Por ejemplo, “一台秤” se refiere a una báscula, “一杆秤” se refiere a una balanza romana. En chino, usamos el mismo carácter “秤”, con dos clasificadores, para reflejar que son dos tipos de “秤”.

个

Entre todos, 个 tiene el uso más amplio.

[ las ventajas de aprender a jugar al ajedrez]

第一点，就是我能比同龄的小孩能坐得住，就是在小的时候吧。毕竟下棋是一个比较静的体育活动，需要全神贯注去思考。第二呢，就是拓展了我的思维，让我的思维更活跃了一点儿。

dì yī diǎn, jiù shì wǒ néng bǐ tóng líng de xiǎo hái néng zuò de zhù, jiù shì zài xiǎo de shí hou ba. bì jìng xià qí shì yí ge bǐ jiào jìng de tǐ yù huó dòng, xū yào quán shén guàn zhù qù sī kǎo. dì èr ne, jiù shì tuò zhǎn le wǒ de sī wéi, ràng wǒ de sī wéi gèng huó yuè le yì diǎnr.

Primero, cuando era pequeño, aguantaba sentado más que los niños de mi edad. En realidad, jugar al ajedrez es una actividad de deporte relativamente tranquila, y necesita pensar con mucha concentración. Segundo, ha ampliado mi pensamiento, y hace que mi pensamiento sea un poco más activo.

Texto: chin02

辆 (liàng , aplicado a vehículos)

[comprar un coche nuevo]

A: 对了，我那个车已经快有将近十年的年头了。

B: 嗯。

A: 我也.....明年想换一辆车，你有什么好的建议么？

A: duì le, wǒ nà ge chē yǐ jīng kuài yǒu jiāng jìn shí

nián de nián tou le。

B: ng。

A: wǒ yě ..... míng nián xiǎng huàn yí **liàng** chē, nǐ yǒu shén me hǎo de jiàn yi me?

A: Ah, aquel coche mío casi va a cumplir 10 años.

B: [asiente] Mmm...

A: Yo también... el año que viene querría cambiarlo por **otro** coche. ¿Tienes buenas sugerencias?

Texto: chin06

**部**(bù, aplicado a novelas, películas, etc.)

**本**(běn, aplicado a libros, etc.)

[la visita a Roma]

不知道你有没有看过一**部**电影叫"天使与魔鬼"？是根据同名的一本**本**小说改编的。这个电影就是在罗马.....嗯.....拍摄的，而且，基本是在罗马的教堂里拍摄的。

bù zhī dào nǐ yǒu méi yǒu kàn guo yí **bù** diàn yǐng jiào "tiān shǐ yǔ mó guǐ"? shì gēn jù tóng míng de yì **běn** xiǎo shuō gǎi biān de。zhè ge diàn yǐng jiù shì zài luó mǎ ..... ng ..... pāi shè de, ér qiě, jī běn shì zài luó mǎ de jiào táng lǐ pāi shè de 。

No sé si has visto **una** película llamada *Ángeles y demonios*. Está basada en **una** novela del mismo nombre. Esta película fue rodada en Roma, además, básicamente en la catedral de Roma.

Texto: chin23

**层**(céng, aplicado a capas, edificios, etc.)

[cómo cocinar la carne]

你下次可以试一下，用小火，把这个火开得越小越好。这样，让那个汤的上面能够形成一**层**完整的油膜。

nǐ xià cì kě yǐ shì yí xià, yòng xiǎo huǒ, bǎ zhè ge huǒ kāi de yuè xiǎo yuè hǎo。zhè yang, ràng nà ge tāng de shàng mian néng gòu xíng chéng yì **céng** wán zhěng de yóu mó。

La próxima vez puedes probar a hacerlo con fuego lento, cuanto más lento, mejor. Así, sobre la superficie de la sopa se formará **una** telilla de grasa.

Texto: chin23

**张**(zhāng, aplicado a papeles)

[mandar postales a los familiares]

那太有纪念意义了, 我一定得给我们家七大姑八大姨都寄一张。

nà tài yǒu jì nián yì yì le, wǒ yí dìng děi gěi wǒ men jiā qī dà gū bā dà yí dōu jì yì **zhāng**。

Es un recuerdo muy bonito. Sin duda, mandaré **una** postal a cada uno de los parientes de nuestra familia [como un recuerdo del viaje].

Texto: chin08

**场**(chǎng, aplicado a espectáculos, juegos, etc.)

[Cómo pasa la Navidad la gente de diferentes países]

而美国人几乎都有带着全家去看一场话剧的爱好。

ér měi guó rén jī hū dōu yǒu dài zhe quán jiā qù kàn yì **chǎng** huà jù de ài hao。

Mientras a casi todos los estadounidenses les gusta llevar a toda la familia a ver **una** obra de teatro.

Texto: chme02

**次**(cì, aplicado a número de veces de cierta acción)

[Para el aprendizaje de idiomas hay que hacer repasos.]

但是第二次课呢, 又遇到了。哎, 有人问, 这是什么? 又不会读了。 我又说了一遍。

dàn shì dì èr **cì** kè ne, yòu yù dào le。 ai, yǒu rén wèn, zhè shì shén me? yòu bú huì dú le。 wǒ yòu shuō le yí biàn。

Pero, **la segunda** clase lo encontramos otra vez. Alguien preguntó qué era esto. No sabía como pronunciarlo. Lo dije otra vez.

Texto: chfo02

**件(jiàn, aplicado a vestidos, asuntos, etc.)**

[plumas]

冬天到了，送给你的朋友和家人一件儿羽绒服，想必一定可以温暖他们的心吧。

dōng tiān dào le, sòng gěi nǐ de péng you hé jiā rén yí **jiàn** r yǔ róng fú, xiǎng bì yí dìng kě yǐ wēn nuǎn tā men de xīn ba.

Ya viene el invierno. Regala a tus amigos y familiares **un** [abrigo de] plumas. Se piensa que podrá calentar su corazón.

Texto: chme01

**份(fèn, aplicado a porciones, periódicos y documentos, etc.)**

[en un currículum]

一份简历做好之后，你还可以去咨询老师以及身边有经验的同学，让他们给你提些参考意见。

yí **fèn** jiǎn lì zuò hǎo zhī hòu, nǐ hái kě yǐ qù zī xún lǎo shī yǐ jí shēn biān yǒu jīng yàn de tóng xué, ràng tā men gěi nǐ tí xiē cān kǎo yì jiàn.

Cuando termines **un** currículum, puedes consultar a los profesores y compañeros cercanos que tienen experiencia para que ellos te den su opinión de referencia.

Texto: chme08

**棵(kē, aplicado a plantas, etc.)**

[exponer el edredón al sol]

如果把被子晒在两棵树中间挂的绳子上，不仅影响校容，而且呀，还会毁坏花草树木。

rú guǒ bǎ bèi zi shài zài liǎng **kē** shù zhōng jiān guà de shéng zi shàng, bù jǐn yǐng xiǎng xiào róng, ér qiě ya, hái huì huǐ huài huā cǎo shù mu.

Si el edredón se expone al sol, colgado de la cuerda entre **dos** árboles, no sólo afecta a la belleza de la universidad, sino que también destruirá flores, hierbas y árboles.

Texto: chme16

### **Clasificador indefinido**

Los más usuales son 点儿 y 些. Sólo se puede añadir 一 delante de los dos. La cantidad que representa 一些 es más que 一点儿.

#### **一些(unos)**

[hablar de la Discoteca Joy de Madrid]

因为那天就是他要让买门票，但是就是那个我们有一些朋友不愿意买那个门票，然后就没有去。

yīn wéi nà tiān jiù shì tā yào ràng mǎi mén piào,  
dàn shì jiù shì nà ge wǒ men yǒu **yì xiē** péng you bù yuàn  
yì mǎi nà ge mén piào, rán hòu jiù méi yǒu qù.

Porque aquel día nos pidió que compráramos entradas,  
pero teníamos **unos** amigos que no querían comprarlas.  
Luego no fuimos.

Texto: chin01

#### **一点(un poco)**

[sobre comer carne en general]

原来你也喜欢吃肉啊！没看出来。但是，你可以自己在家里煮一点肉吃啊？或者是在炒菜的时候，多放点肉，都是……都是好办法。

yuán lái nǐ yě xǐ huān chī ròu a! méi kàn chū lai.  
dàn shì, nǐ kě yǐ zì jǐ zài jiǎ lǐ zhǔ **yì diǎn** ròu chī  
a? huò zhě shì zài chǎo cài de shí hou, duō fàng diǎn  
ròu, dōu shì…… dōu shì hǎo bàn fa.

¡También te gusta comer carne! No se nota. Pero tú misma puedes cocer **un poco de** carne en casa. O, cuando fríes verduras, echa un poco más de carne, todos son buenos métodos.

Texto: chin07

### **Fecha y hora**

En chino la hora y la fecha siguen el mismo orden: de mayor a menor, de años a días y de horas a segundos.

## Fecha

[fecha de los partidos de fútbol de los Juegos Olímpicos de Beijing]

比赛日期为 **2008 年 8 月 6 日至 23 日**。

bǐ sài rì qī wéi èr líng líng bā nián bā yuè liù  
rì zhì èr shí sān rì。

La fecha de los partidos es **del 6 de agosto al 23 de agosto del año 2008**.

Texto: chme06

## Hora

[hablando sobre ir a Correos en España por la mañana para mandar postales]

因为我来了以后，看到西班牙的工作方式一般都是在下午**两点左右**下班。所以我一般如果要办事情，都在上午**十点**或者**十一点**之间把我的事情办好。

yīn wéi wǒ lái le yǐ hòu, kàn dào xī bān yá de gōng  
zuò fāng shì yì bān dōu shì zài xià wǔ **liǎng diǎn zuǒ  
yòu** xià bān。suǒ yǐ wǒ yì bān rú guǒ yào bàn shì qing,  
dōu zài shàng wǔ **shí diǎn** huò zhě **shí yī diǎn** zhī jiān  
bǎ wǒ de shì qing bàn hǎo。

Porque después de venir aquí, veía que en general, la forma de trabajo de España era terminar el trabajo **más o menos a las dos** de la tarde. Por eso, si quiero hacer algo, suelo hacerlo **entre las diez y las once** de la mañana.

Texto: chin08

Aquí **左右**( más o menos, aproximadamente) siempre se coloca detrás del tiempo.

## Fecha y hora

[el lanzamiento del satélite Chang'e 1 hacia la luna]

**10 月 24 日晚 6 点 5 分**，我国嫦娥一号卫星在西昌卫星发射中心发射升空。在这个全国人民瞩目的时刻，我校近千名学子纷纷自发地聚集在大屏幕前观看了"嫦娥一号"奔月的电视直播。

shí yuè èr shí sì rì wǎn liù diǎn wǔ fēn, wǒ guó

cháng é yī hào wèi xīng zài xī chāng wèi xīng fā shè  
zhōng xīn fā shè shēng kōng。 zài zhè ge quán guó rén  
mín zhǔ mù de shí kè, wǒ xiào jìn qiān míng xué zǐ fēn  
fēn zì fā de jù jí zài dà píng mù qián guān kàn le “  
cháng é yī hào ” bēn yuè de diàn shì zhí bō。

**A las seis y cinco del día 24 de octubre**, el satélite  
explorador lunar Chang’ e 1 se lanzó al cielo en el  
Centro de Lanzamiento de Satélites Xichang. En este  
momento que llamó la atención de todo el país, casi mil  
estudiantes de la universidad se reunieron  
voluntariamente delante de la pantalla grande para ver  
el lanzamiento de Chang’ e 1 hacia la luna en la  
televisión en directo.

Texto: chme16

### **Media hora**

[entrada gratis para ver una representación]

En chino, para expresar “media hora”, hay dos formas. Una es usar  
“半(media)”. Otra es decir directamente “30 分 (30 minutos)”.

周日晚上**六点半**明德厅见!免票入场!

zhōu rì wǎn shang **liù diǎn bàn** míng dé tīng jiàn!  
miǎn piào rù chǎng!

¡Nos veremos **a las seis y media** el domingo por la noche  
en el salón Ming De! ¡Entrada gratis!

Texto: chme03

[participar en una gymkana de universitarios]

下午 **2 点 30 分**, 我校同学以饱满的精神, 出现在人大的操  
场上。市委教育工委高校奥运工作办公室主任宣布首都大学生趣  
味奥运挑战赛正式启动。

xià wǔ **liǎng diǎn sān shí fēn**, wǒ xiào tóng xué yǐ  
bǎo mǎn de jīng shen, chū xiàn zài rén dà de cāo chǎng  
shang。 shì wěi jiào yù gōng wěi gāo xiào ào yùn gōng  
zuò bàn gōng shì zhǔ rèn xuān bù shǒu dū dà xué shēng  
qù wèi ào yùn tiǎo zhàn sài zhèng shì qǐ dòng。

**A las 2 y media** de la tarde, llenos de vitalidad, los



estudiantes de nuestra universidad aparecieron en el campo de la Universidad del Pueblo. El director de la Oficina del Trabajo Universitario de los Juegos Olímpicos del Comité de Enseñanza del Comité Municipal anunció la puesta en marcha de una gymkana de los Juegos Olímpicos para los universitarios de la capital.

Texto: chme15

En chino, las palabras que indican tiempo suelen colocarse delante del predicado y especifican cuándo ocurre la acción del verbo. En este sentido, la colocación de las palabras de tiempo del español es más flexible, por ejemplo, se puede decir “A las seis y media del día 25 de mayo” o “El 25 de mayo a las seis y media”. La posición sintáctica de la palabra depende del énfasis del hablante.

### **Locuciones del chino 成语 (Chengyu)**

*Chengyu* es una peculiaridad de las locuciones del chino y equivale a los refranes o modismos del español. Se caracteriza por su significado completo, formato fijo y expresión concisa. Procede de obras literarias clásicas, libros históricos, lenguaje popular, etc. La gran mayoría de los *chengyu* están compuestos de cuatro caracteres, hay muy pocos compuestos por menos. En cuanto al número de caracteres de locuciones del español, no hay un principio básico o una costumbre como tal. Es más flexible.

[la forma de trabajo de un chico]

工作中的他同样也**一丝不苟**，认真对待每一件事。

gōng zuò zhōng de tā tóng yàng yě **yī sī bù gǒu**,  
rèn zhēn duì dài měi yí jiàn shì。

Hace su trabajo **a conciencia y con mucho cuidado**,  
tratando cada asunto con atención.

Texto: chme14

[hablando sobre juegos en internet]

但是我觉得，那些网游都**大同小异**。

dàn shì wǒ jué de, nèi xiē wǎng yóu dōu **dà tóng xiǎo yì**。

Pero, me parece que aquellos juegos en internet son **idénticos en lo fundamental, pero diferentes en los detalles**.

Texto: chin12

[cómo hacer ejercicios físicos en otoño]

秋季锻炼和其他季节锻炼一样，运动量应由小到大、**循序渐进**。

qiū jì duàn liàn hé qí tā jì jié duàn liàn yí yàng,  
yùn dòng liàng yìng yóu xiǎo dào dà、**xún xù jiàn jìn**。  
Igual como en otras estaciones, la intensidad de los  
ejercicios físicos en otoño tiene que ir de menos a más,  
**gradualmente**.

Texto: chme03

[la dieta china]

咱们中国的饮食真是**博大精深**，连每一个蔬菜对应的什么功效，都能说得**一清二楚**。

zán men zhōng guó de yǐn shí zhēn shì **bó dà jīng shēn**,  
lián měi yí gè shū cài duì yìng de shén me gōng xiào,  
dōu néng shuō de **yī qīng èr chǔ**。

La dieta china es realmente **amplia y variada**, incluso el efecto correspondiente de cada verdura se puede explicar **claramente**.

Texto: chin19

[Cómo dividir el tiempo de manera razonable para el estudio]

所以我们一定要合理分配时间，控制好英语学习及专业学习的比重，这样才能**事半功倍**。

suǒ yǐ wǒ men yí dìng yào hé lǐ fēn pèi shí jiān,  
kòng zhì hǎo yīng yǔ xué xí jí zhuān yè xué xí de bǐ zhòng,  
zhè yàng cái néng **shì bàn gōng bèi**。

Por eso, tenemos que dividir el tiempo de manera razonable, equilibrando la proporción entre el estudio de inglés y el estudio de la carrera, de esta manera, podremos **lograr doble resultado con la mitad del esfuerzo**.

Texto: chme13

[adornar el dormitorio con plantas o flores]

噢，我是因为在国内的时候，我妈妈喜欢养这些花花草草，所以出来以后，我觉得，反正我也是住单人间，把这个屋子里弄得**生机勃勃**一点，我自己的心情也好。

o, wǒ shì yīn wéi zài guó nèi de shí hou, wǒ mā ma xǐ huān yǎng zhèi xiē huā huā cǎo cǎo, suǒ yǐ chū lái yǐ hòu, wǒ jué de, fǎn zhèng wǒ yě shì zhù dān rén jiān, bǎ zhè ge wū zi lǐ nòng de **shēng jī bó bó** yì diǎn, wǒ zì jǐ de xīn qíng yě hǎo.

Pues porque cuando estaba en China, a mi madre le gustaba cultivar flores e hierbas. Por eso después de salir de China, me parece que, como vivo en una habitación individual, si puedo llenarla **de vitalidad**, estaré de muy buen humor.

Texto: chin19

## 咱们 y 我们

En español, hay una única palabra para la primera persona del plural “nosotros”. En chino hay dos palabras: “咱们” incluye a los hablantes y los oyentes, mientras que “我们” incluye sólo a los hablantes. En español, cuando es necesario expresar esta diferencia, se explicita posteriormente.

## 咱们 nosotros inclusivo

[jugando a las cartas para comprobar quién juega mejor]

好啊，**咱们**去一较高低。

hǎo a, **zán men** qù yí jiào gāo dī.

Bueno, **vamos** a echar unas partidas de cartas.

Texto: chin02

## 我们 nosotros exclusivo

[programa de televisión *La vida de la moda*]

Hello, 观众朋友们大家好! 欢迎收看本期时尚生活。让我们用时尚点亮你的生活。

Hello, guān zhòng péng yǒu men dà jiā hǎo! huān yíng shōu kàn běn qī shí shàng shēng huó。ràng **wǒ men** yòng shí shàng diǎn liàng nǐ de shēng huó。

¡Hola a todos los espectadores! ¡Bienvenida a nuestro programa *La vida de la moda*! Déjanos (**a nosotros**) hacer brillar tu vida con la moda.

Texto: chme01

## Oración interrogativa

Para convertir una oración enunciativa en una interrogativa, hay diferentes formas.

1. Cambiar la entonación.

1.a. Se puede añadir al final de la oración partículas interrogativas como 吗, 呢, etc.

[jugar juegos en internet con otras personas]

那在这个小团队里, 你能当个小头目吗?

nà zài zhè ge xiǎo tuán duì lǐ, nǐ néng dāng ge xiǎo tóu mù **ma**?

En este equipo pequeño, ¿puedes servir de líder?

Texto: chin01

1.b. Este tipo de oraciones interrogativas puede no llevar partículas interrogativas. En este sentido, es igual como el español.

[el uso de la crema solar para ir a la playa]

那你没买那个防晒霜? 就是抹上?

nà nǐ méi mǎi nà ge fáng shài shuāng? jiù shì mǒ shàng?

¿No te has comprado esa crema solar? ¿Para untártela?

Texto: chin01

2.a. Añadir pronombres interrogativos como 什么, 怎么样, etc. y cambiar la entonación. Puede llevar partículas interrogativas.

A: 是什么牌子呀?

B: 不知道。

A: shì **shén me** pái zi **ya**?

B: bù zhī dào。

A: ¿De **qué** marca es?

B: No sé.

Texto: chin01

2.b. Puede no llevar partículas interrogativas, pero siempre un pronombre interrogativo.

那你现在象棋的水平**怎么样**了?

nà nǐ xiàn zài xiàng qí de shuǐ píng **zě n me yàng**  
le?

¿**Qué tal** es tu nivel de ajedrez ahora?

Texto: chin02

3.a. Oración interrogativa con dos opciones.

A: 兔子叫什么呢?

B: 嗯, 是西语名**还是**中文名 ?

A: 当然是西语名了。

B: 嗯, 它叫 LECHUGA。

A: tù zi jiào shén me ne?

B: ng, shì xī yǔ míng **hái shì** zhōng wén míng?

A: dāng rán shì xī yǔ míng le。

B: ng, tā jiào LECHUGA。

A: ¿Cómo se llama el conejo?

B: Mmm ..., ¿su nombre español **o** chino?

A: Por supuesto que es su nombre español.

B: Mmm ..., se llama Lechuga。

Texto: chin13

3.b. También pueden añadirse partículas interrogativas.

A: 嗯.....这样的话,你整体在西班牙,就是过了这三个月,你认为 [chasquido] 就她给你留下一个怎样的印象呢?

B: 是说西班牙**还是**这些欧洲国家**啊**?

A: 嗯.....先西班牙吧!

A: ng ..... zhè yang de hua, nǐ zhěng tǐ zài xī bān yá, jiù shì guò le zhè sān ge yuè, nǐ rèn wéi [chasquido] jiù tā gěi nǐ liú xià yí ge zěn yàng de yìn xiàng ne?

B: shì shuō xī bān yá **hái shì** zhèi xiē ōu zhōu guó jiā **a**?

A: ng ..... xiān xī bān yá ba!

A: Mmm... así, has estado en total tres meses en España. Según tú, ¿qué impresión te ha dejado?

B: ¿Te refieres a España **o** a los países europeos?

A: Mmm ...primero España.

Texto: chin22

## Partículas estructurales 的、地、得

### 的+sustantivo

**的** suele usarse detrás de sustantivos o adjetivos. La combinación del modificador adjetival que le precede y el sintagma nominal que le sigue puede expresar las características o la relación de posesión, etc.

[presentando el protocolo de recibir y hacer llamadas]

在这期**的**礼仪版块中,由我来为大家介绍一下接打电话**的**礼仪。

zài zhè qī **de** lǐ yí bǎn kuài zhōng, yóu wǒ lái wèi dà jiā jiè shào yí xià jiē dǎ diàn huà **de** lǐ yí.

En la parte de protocolo **de** este programa, os voy a presentar a todos el protocolo **de** recibir y hacer llamadas.

Texto: chme19

### adjetivo +地+ verbo

En general, es palabra símbolo del modificador adverbial en chino. Suele colocarse detrás de un adjetivo u otras palabras, formando con

ellos un modificador adverbial, y delante de algún verbo.

[En la antigüedad, Da Yu estudia por qué su padre no puede controlar la inundación.]

他一想啊，他的爸爸辛辛苦苦地干了九年，都没有任何成绩，最后还被罢免了，别人还都埋怨他。他就知道，他爸爸的方法，一定是不对的。

tā yì xiǎng a, tā de bà ba xīn xīn kǔ kǔ **de** gàn le jiǔ nián, dōu méi yǒu rèn hé chéng jì, zuì hòu hái bèi bà miǎn le, bié rén hái dōu mǎn yuàn tā. tā jiù zhī dao, tā bà ba de fāng fǎ, yí dìng shì bú duì de.

Pensaba que su padre había trabajado duramente durante nueve años sin lograr éxito alguno, al final, había sido depuesto, y todo el mundo le echaba la culpa. Sabía seguro que el método de su padre no era correcto.

Texto: chfo04

### Verbo+得+adjetivo

Es un sintagma verbal y complemento. El complemento indica la posibilidad y la situación de la acción del verbo. Para saber el grado a que llega el verbo, se puede hacer la pregunta con “怎么样”. Por ejemplo, “我们听得不太明白”, para saber el grado de “听”, se puede hacer la pregunta con “听得怎么样?”. En muchos casos, al traducirse al español, no aparece.

[Los estudiantes extranjeros entienden mejor la clase.]

我们班上有一个突尼斯的小伙子。虽然他说的西语，我们听得不明白，但是老师能听懂。所以上课的时候，他经常和老师讨论问题，我觉得他们比我们接收得更好。

wǒ men bān shàng yǒu yí gè tū ní sī de xiǎo huǒ zi. suī rán tā shuō de xī yǔ, wǒ men tīng **de** bú tài míng bai, dàn shì lǎo shī néng tīng dǒng. suǒ yǐ shàng kè de shí hou, tā jīng cháng hé lǎo shī tāo lùn wèn tí, wǒ jué de tā men bǐ wǒ men jiē shōu **de** gèng hǎo.

En nuestro grupo, hay un chico de Túnez. Aunque no podemos entender bien el español hablado por él, el profesor puede entenderle. Por eso, en clase, discute

a menudo con el profesor sobre preguntas. Me parece que ellos entienden mejor que nosotros.

Texto: chin23

### Suplemento direccional compuesto

La estructura gramatical “predicado + suplemento direccional” siempre ha constituido un punto difícil para los extranjeros porque en muchas lenguas no la hay. Su función consiste en indicar hacia dónde se realiza o se ha realizado la acción del verbo, por ejemplo: 上来, 上去, 下来, 下去, 进来, 进去, 出来, 出去, 回来, 回去, 过来, 过去, 起来, etc. En general, la dirección del desarrollo de la acción, depende de la posición del hablante: hacia él o desde él.

[el horario de actividades de una mascota]

这倒不是的。因为上次你**过来**的时候，是下午的时候。然后我们众所周知，兔子跟我们人类的时间差是很大的。一般它的活动时间是在晚上十点之后。

zhè dào bú shì de。yīn wéi shàng cì nǐ **guò lai** de shí hou , shì xià wǔ de shí hou。rán hòu wǒ men zhòng suǒ zhōu zhī, tù zi gēn wǒ men rén lèi de shí jiān chā shì hěn dà de。yì bān tā de huó dòng shí jiān shì zài wǎn shàng shí diǎn zhī hòu。

No es así. Porque la vez pasada cuando **viniste**, era la tarde. Como todos sabemos, los conejos tienen un horario muy diferente del nuestro. Generalmente, su horario de actividades es después de las diez de la noche.

Texto: chin13

[el orden de las reformas de la casa]

嗯，比如说，我必须要在它地板装好之后、瓷砖铺安装好之后，我才能安排这个送橱柜的人**过去**给我安装橱柜。这个所有的工序你都要安装好了。

ng, bǐ rú shuō, wǒ bì xū yào zài tā dì bǎn zhuāng hǎo zhī hòu、cí zhuān pū ān zhuāng hǎo zhī hòu, wǒ cái néng ān pái zhè ge sòng chú guì de rén **guò qu** gěi wǒ



ān zhuāng chū guì。zhè ge suǒ yǒu de gōng xù nǐ dōu yào  
ān zhuāng hǎo le。

Por ejemplo, tengo que mandar al hombre a llevar e  
instalar la alacena **allí**, y eso sólo se puede hacer  
después de instalarse el parquet y los azulejos. Tienes  
que preparar bien el orden de las reformas.

Texto: chin17

En los dos ejemplos arriba mencionados, uno usa “过来”, el otro usa  
“过去”. “过” significa moverse en el mismo nivel horizontal del  
espacio. La diferencia entre “来” y “去” es: “来” indica moverse  
hacia el hablante, y “去”, alejarse del hablante, moviéndose hacia el  
otro lado. En el primer ejemplo, “过来” se refiere a venir al piso del  
interlocutor, mientras en el segundo, “过去”, se refiere a que el  
hablante manda al obrero al piso de un cliente.

[aprender a jugar al ajedrez]

那这个国际象棋学**起来**难吗？

nà zhè ge guó jì xiàng qí xué **qǐ lai** nán ma?

¿Es difícil aprender el ajedrez?

Texto: chin02

En la traducción al español a veces no se incluye el suplemento  
direccional, que en este caso indica “el inicio y el seguimiento” de  
aprender ajedrez.

[Después de cambiar el estilo de vestirse, la chica salió a una  
discoteca.]

A: 她自己那天是不是自己也特别开心？

B: 她当然自己特别开心了。然后**出去**，所有的男生都跟她说话。

A: nà tā zì jǐ nà tiān shì bú shì zì jǐ yě tè bié kāi  
xīn?

B: tā dāng rán zì jǐ tè bié kāi xīn le。rán hou **chū qu**,  
suǒ yǒu de nán shēng dōu gēn tā shuō huà。

A: Aquel día, ¿estaba muy contenta?

B: Por supuesto estaba muy contenta. Y cuando **salimos**,

todos los chicos querrían hablar con ella.

Texto: chin01

[la afición de pintar]

然后就是如果是觉得什么事情特别好玩，也会把它画**下来**。

rán hou jiù shì rú guǒ shì jué de shén me shì qīng  
tè bié hǎo wán, yě huì bǎ tā huà **xià lai**。

Y si me parece que algo es muy interesante, también lo  
pintaré.

Texto: chin01

En este caso, “下来” tampoco se traduce al español e indica como  
“guardar algo”.

[juegos en internet]

特别好玩儿。我天天心心念念都想着**回去**打魔兽。

tè bié hǎo wánr. wǒ tiān tiān xīn xīn niàn niàn dōu  
xiǎng zhe **huí qu** dǎ mó shòu。

Es muy interesante. Todos los días pienso en **volver**  
para jugar al [juego] Mo Shou.

Texto: chin01

**Tono:**

El chino *putonghua* tiene cuatro tonos fundamentales. Una misma sílaba puede tener distintos tonos que denotan diferentes significados. Por ejemplo, para la sílaba *ma*, tiene cuatro tonos que son mā, má, mǎ y mà, y sus cuatro significados nada tienen que ver entre sí.

**El primer tono**

**mā** (*madre*) 妈

[hablar del amor del pasado]

嗯，其实我也是一个挺怀旧的人。但是，我**妈(mā)**就常劝我说，就是人不要这么怀旧，不要老看着过去，要往前看。

ng, qí shí wǒ yě shì yí ge tǐng huái jiù de rén.  
dàn shì, wǒ **mā** jiù cháng quàn wǒ shuō, jiù shì rén bú  
yào zhè me huái jiù, bú yào lǎo kàn zhe guò qù, yào wǎng  
qián kàn。

Pues, de hecho, también soy una persona que extraña mucho el pasado. Pero mi **madre** siempre me aconsejaba diciendo que la gente no debería extrañar tanto el pasado ni mirar siempre hacia el pasado, sino que tendría que mirar hacia el futuro.

Texto: chin14

### El segundo tono

**má** (complicado) 麻(junto con 烦)

[hablando sobre el envío de un paquete por correo]

啊, 这么**麻**(má)烦? 我以前可从来都没听说过。

a, zhè me **má** fan? wǒ yǐ qián kě cóng lái dōu méi tīng shuō guo.

¿Ah, tan **complicado**? Antes nunca había oído hablar de eso.

Texto: chin08

### El tercer tono

**mǎ** (número) 码

[principios del partido de fútbol de los Juegos Olímpicos]

队员上衣背后、前胸和短裤前应有号码**码**(mǎ)。队长要佩戴袖标, 上场队员必须穿足球鞋。

duì yuán shàng yī bèi hòu、qián xiōng hé duǎn kù qián yīng yǒu hào **mǎ**。duì zhǎng yào pèi dài xiù biāo, shàng chǎng duì yuán bì xū chuān zú qiú xié。

En la espalda, el pecho y los pantalones cortos de los jugadores, hay que poner el **número**. El jefe del equipo debe llevar brazalete, y los jugadores en el campo deben llevar zapatos de fútbol.

Texto: chme06

### El cuarto tono

**mà** (criticar) 骂

[aprender música desde pequeño]

我觉得你实在太厉害了。因为我小时候也被我妈逼着学过手风琴, 但我从来都不喜欢练。每次都是要被老师**骂**(mà)很长时间。

wǒ jué de nǐ shí zài tài lì hai le。 yīn wéi wǒ xiǎo shí hou yě bèi wǒ mā bī zhe xué guò shǒu fēng qín, dàn wǒ cóng lái dōu bù xǐ huān liǎn。 měi cì dōu shì yào bèi lǎo shī **mà** hěn cháng shí jiān。

Me parece que eres muy fuerte. Es porque de pequeña, obligada por mi madre, también aprendía acordeón, aunque jamás me gustó ensayar. El maestro me **criticaba** a menudo cada vez que lo tocaba.

Texto: chin14

Lo mismo pasa con las palabras que tienen dos sílabas.

### **jiào shī** (*profesor*) 教师

[trabajo del decano de la facultad de inglés]

他亲身参与每一部分建设, 付出了巨大心血。他与**教师**(jiào shī)们召开多次研讨会, 聆听**教师**们的想法。

tā qīn shēn cān yù měi yí bù fēn jiàn she, fù chū le jù dà xīn xuè。 tā yǔ **jiào shī** men zhào kāi duō cì yán tǎo huì, líng tīng **jiào shī** men de xiǎng fǎ。

Se implicaba personalmente en el progreso de cada parte, empeñándose con mucho esfuerzo. Convocaba muchas veces seminarios con los **profesores**, escuchando **sus** opiniones.

Texto: chme13

### **jiào shì** (*aula*) 教室

[aprender a hacer jiaozi]

在面粉横飞的**教室**(jiào shì)中, 弥漫着和谐温暖的气氛。

zài miàn fěn héng fēi de **jiào shì** zhōng, mí màn zhe hé xié wēn nuǎn de qì fēn。

En el **aula** donde vuela el polvo de trigo, se extendía el ambiente armonioso.

Texto: chme12

Nota: Hay palabras que tienen las mismas sílabas y tonos. Para distinguirlas, hay que entenderlas según el contexto:

[carácter de los animales]

因为我小时候有一次下大雨(dà yǔ), 住平房, 然后我爸就在那个屋檐底下, 就捡了一只小麻雀回来。但就养到第二天, 想等它好点把它放走, 结果它就死了。好像是自己咬断舌头自尽了, 因为麻雀就特别有性格。

yīn wéi wǒ xiǎo shí hou yǒu yí cì xià dà yǔ, zhù píng fáng, rán hòu wǒ bà jiù cóng nà ge wū yán dī xia, jiù jiǎn le yí zhī xiǎo má què huí lai。dàn jiù yǎng dào dì èr tiān, xiǎng děng tā hǎo diǎn bǎ tā fàng zǒu, jié guǒ tā jiù sǐ le。hǎo xiàng shì zì jǐ yǎo duàn shé tou zì jìn le, yīn wéi má què jiù tè bié yǒu xìng gé 。  
Porque recuerdo que cuando era niña, en una ocasión **llovía fuertemente**. Vivía en una casa que sólo tenía una planta. Mi padre recogió un pequeño gorrión debajo del tejado. Pero sólo lo criamos hasta el día siguiente. Esperábamos soltarlo cuando mejorara un poco. Resulta que murió. Parece ser que se mordió la lengua causándose su propia muerte porque los gorriones tienen mucho carácter.

Texto: chin13

[leyenda de Dayu]

下面再给大家讲一个.....嗯.....中国人都了解的一个.....一个故事, 叫做大禹(dà yǔ)治水。

xià mian zài gěi dà jiā jiǎng yí ge ng zhōng guó rén dōu liǎo jiě de yí ge yí ge gù shi, jiào zuo dà yǔ zhì shuǐ。

A continuación, voy a contarles un cuento que todos los chinos conocen, llamado el control de inundación de **Da Yu**.

Texto: chfo04

Aquí el primer *dà yǔ* significa “lluvia grande” y el segundo es un nombre de un chino de la antigüedad. En ese caso los pictogramas “大禹” asumen un valor meramente fonográfico aunque, en este caso, la significación de cada pictograma podría también querer transmitir

algo acerca de la personalidad de ese hombre, etc. Cuando escuchamos estas dos secuencias de sonidos, sólo podemos distinguirlas en el contexto. En la escritura, sin embargo, sí se ve que son distintos caracteres por mucho que se pronuncien de la misma forma.

### **Cambio de tono**

**Tono ligero:** Cuando algunas sílabas están detrás de otras formando con ellas algún concepto o nombre unitario, pierden sus tonos originales. Se pronuncia con menos fuerza. En la escritura, no lleva signo de tono. Por ejemplo:

[cómo vivía la gente en la antigüedad]

嗯，外出去打渔啊、或者是打猎啊，是碰碰**运气**。

ng, wài chū qù dǎ yú a, huò zhě shì dǎ liè a, shì pèng pèng **yùn qi**。

Pues el salir a pescar o cazar, es para tentar a la **suerte**.

Texto: chfo05

En vez de pronunciarse *yùmqì*, 运气 se pronuncia como *yùnqi*.

Los casos fundamentales en que se verifica esta **neutralización de tono** son:

1. partícula interrogativa “吗、呢、啊、吧”, etc.

[juegos en internet]

那你怎么办？你这边儿打魔兽，那边儿你还开着 MSN 或者是这个 QQ，然后随时跟人家交流是**吗**？

nà nǐ zěn me bàn? nǐ zhè biānr dǎ mó shòu, nà biānr nǐ hái kāi zhe MSN huò zhě shì zhè ge QQ, rán hòu suí shí gēn rén jiā jiāo liú shì **ma**?

Entonces, ¿qué haces? Juegas a Moshou aquí y allí dejas abierto tu MSN o QQ, y luego haces intercambio con otros, ¿es así?

Texto: chin01

[ir juntas al supermercado a hacer la compra]

A: 行，那咱们走吧！

B: 走吧！

A: 嗯！

A: xíng, nà zán men zǒu ba!

B: zǒu ba!

A: ng!

A: Vale, ¡vamos!

B: ¡Venga, vamos!

A: ¡Venga!

Texto: chin19

2. partícula “着、了、过、的、地、得、们”, etc.

[hablando de recibir un paquete en España]

啊？你不是刚到西班牙，怎么包裹就来了？

a? nǐ bú shì gāng dào xī bān yá, zěn me bāo guǒ jiù lái le?

Ah, acabas de llegar a España, ¿ha llegado ya tu paquete?

Texto: chin08

[la vida de la moda]

Hello, 观众朋友们大家好！欢迎收看本期时尚生活。让我们用时尚点亮你的生活。

Hello, guān zhòng péng yǒu men dà jiā hǎo! huān yíng shōu kàn běn qī shí shàng shēng huó。ràng wǒ men yòng shí shàng diǎn liàng nǐ de shēng huó。

¡Hola a todos los espectadores! ¡Bienvenida a nuestro programa “*Vida de la moda*”! Déjanos hacer brillar tu vida con la moda.

Texto: chme01

3. sufijo de sustantivo: “子、头”, etc.

[plantas en el dormitorio]

而且你没发现我宿舍的柜子是绿色的吗？我在桌子上摆一盆仙人球儿，还挺配合我这个宿舍的整体风格呢！

ér qiě nǐ méi fā xiàn wǒ sù shě de guì **zì** shì lǜ sè de ma? wǒ zài zhuō zi shàng bǎi yī pén xiān rén qiú, hái tǐng pèi hé wǒ zhè ge sù shě de zhěng tǐ fēng gé ne! Además, ¿no has notado que el armario de mi dormitorio es verde? He puesto un cactus sobre la mesa, que pega con el estilo completo de mi dormitorio.

Texto: chin19

[sacar fotos durante el viaje a Europa]

嗯，在城市里面自然是拍一些城市的建筑，或者是博物馆里头的艺术品。然后在一些风景比较美丽的地区，拍拍大海啊，拍拍天空，这些也挺好的。

ng, zài chéng shì lǐ miàn zì rán shì pāi yī xiē chéng shì de jiàn zhù, huò zhě shì bó wù guǎn lǐ **tóu** de yì shù pǐn。 rán hòu zài yī xiē fēng jǐng bǐ jiào měi lì de dì qū, pāi pāi dà hǎi a, pāi pāi tiān kōng, zhèi xiē yě tǐng hǎo de。

En la ciudad, sin duda voy a sacar fotos de las obras de arquitectura o de piezas de arte en el museo. Y luego en los lugares pintorescos, voy a sacar fotos del mar, del cielo, que también son muy buenas.

Texto: chin22

4. palabras de lugar: “上、下”, etc.

[plumas en invierno]

那现在的天气是一天比一天冷了。同学们身上的外衣也渐渐地由单衣、棉服转向了羽绒服。一提到羽绒服，我想同学们印象大多都是臃肿和笨重。

nà xiàn zài de tiān qì shì yì tiān bǐ yì tiān lěng le。 tóng xué men shēn shàng de wài yī yě jiàn jiàn de yóu dān yī、mián fú zhuǎn xiàng le yǔ róng fú。 Yì tí dào yǔ róng fú, wǒ xiǎng tóng xué men yìn xiàng dà duō dōu shì yōng zhǒng hé bèn zhong。



Bueno, ahora hace cada día más frío. La ropa que llevan los compañeros va cambiando gradualmente de vestido sin forro o vestido de algodón a plumas. Al referirse a plumas, imagino que a los compañeros os da la impresión de ser gordo y pesado.

Texto: chme01

5. La última sílaba de un nombre o forma verbal duplicados.

Por ejemplo:      说说   想想   弟弟   奶奶   谈谈   跳跳

6. La última sílaba de un verbo direccional como 出来, 进去, 站起来, etc.

[clase de chino]

如果你能够把课文很熟练地说**出来**, 那么, 其中的这个语言点, 你也就掌握了。

rú guǒ nǐ néng gòu bǎ kè wén hěn shú liàn de shuō **chū lai**, nà me, qí zhōng de zhè ge yǔ yán diǎn, nǐ yě jiù zhǎng wò le.

Si puedes decir el texto de manera fluída, entonces, entre otros, este punto gramatical ya lo habrás dominado.

Texto: chfo01

7. la última sílaba de ciertas palabras usuales que tienen dos sílabas.

[pasas y uvas]

我吃着**葡萄**的时候, 就想起来了咱们中国的葡萄干儿。葡萄干儿可比**葡萄**要甜多了。不过论营养价值, 估计还是新鲜的**葡萄**比较高呀!

wǒ chī zhe **pú tao** de shí hou, jiù xiǎng qǐ lái le zán men zhōng guó de pú tao gānr. pú tao gānr kě bǐ **pú tao** yào tián duō le. bù guò lùn yíng yǎng jià zhí, gū ji hái shì xīn xiān de **pú tao** bǐ jiào gāo ya!

Cuando como **uvas**, recuerdo las pasas de China. Las pasas son más dulces que las **uvas**. Pero en cuanto al valor

alimenticio, calculo que las **uvas** frescas tienen más.

Texto: chin19

不

Cuando la sílaba *bù* está delante de otra sílaba del cuarto tono, se pronuncia en segundo tono. Por ejemplo:

[jugar al ajedrez]

有一个棋魂，哎，**不对**，那个**不是**讲下棋的事，下象棋的事情，是在讲下围棋的事情。

yǒu yí ge qí hún, ai, **bú** duì, nà ge **bú** shì jiǎng xià qí de shì, xià xiàng qí de shì qing, shì zài jiǎng xià wéi qí de shì qing.

Existe un espíritu ajedrecístico. Huy, **no**, eso **no** es para hablar de ajedrez, de asuntos de ajedrez, sino de Go.

Texto: chin02

En vez de pronunciarse *bù duì* y *bù shì*, se pronuncia como *bú duì* y *bú shì*.

一

Cuando está en una sílaba simple o al final de una palabra, se pronuncia como *yī*.

[pasar la Navidad]

嗯，其实也没什么，毕竟自己是大三的学生吗？然后考试也特别多，有十一门专业课。嗯，然后……其实主要还是准备考试，但是到圣诞节，也到元旦了，然后，应该是跟朋友一起过吧！

ng, qí shí yě méi shén me, bì jìng zì jǐ shì dà sān de xué shēng ma? rán hòu kǎo shì yě tè bié duō, yǒu shí **yī** mén zhuān yè kè. ng, rán hòu……qí shí zhǔ yào hái shì zhǔn bèi kǎo shì, dàn shì dào shèng dàn jié, yě dào yuán dàn le, rán hòu, yīng gāi shì gēn péng you yì qǐ guò ba!

Pues, en realidad, no hay nada especial. A fin de cuentas,

soy estudiante de tercer curso. Luego tengo muchos exámenes. Hay 11 asignaturas de la carrera. Luego... En realidad, tengo que prepararme para el examen. Pero ya llegamos a la Navidad y al Año Nuevo. Después, voy a pasarlo con amigos.

Texto: chme02

Cuando la sílaba *yī*(一) está delante de otra sílaba del cuarto tono, se pronuncia en segundo tono(*yí*).

[cómo se pone la ropa en clase de deporte]

上体育课或外出锻炼时，应该多穿件宽松舒适的外套。等准备活动做完，或锻炼一会儿，身体发热后，再脱下外衣。免得室内外温差太大，身体不适应而着凉感冒。

shàng tǐ yù kè huò wài chū duàn liàn shí, yīng gāi duō chuān jiàn kuān sōng shū shì de wài tào。děng zhǔn bèi huó dòng zuò wán, huò duàn liàn **yí** huìr, shēn tǐ fā rè hòu, zài tuō xià wài yī。miǎn de shì nèi wài wēn chā tài dà, shēn tǐ bú shì yīng ér zháo liáng gǎn mào。  
En la clase de deporte o cuando hacemos ejercicios fuera, hay que llevar abrigos anchos y cómodos. Cuando terminamos las actividades preparativas o practicamos un poco, después de calentar el cuerpo, nos quitamos el abrigo, evitando que debido a la gran diferencia de temperatura entre la habitación y fuera, el cuerpo no se adapte y coja el frío y la gripe.

Texto: chme03

Cuando la sílaba *yī*(一) está delante de otra sílaba del primer, segundo o tercer tono, se pronuncia en cuarto tono(*yì*).

[la construcción de la Gran Muralla]

一直修到哪儿呢？修到这个现在的蒙古共和国，还有内蒙古自治区，一直修到新疆，修到黑龙江。几乎有两万里长，也就是几乎有一万公里长。

**yì zhí** xiū dào nǎr ne? xiū dào zhe ge xiàn zài de

měng gǔ gòng hé guó, hái yǒu nèi měng gǔ zì zhì qū,  
**yì zhí** xiū dào xīn jiāng, xiū dào hēi lóng jiāng。jī  
hū yǒu liǎng wàn lǐ cháng, yě jiù shì jī hū yǒu yí wàn  
gōng lǐ cháng。

¿Hasta dónde se construye? Hasta la actual República de Mongolia y la región autónoma de Mongolia Interior, hasta Xin Jiang, hasta Hei Longjiang. Casi hay 20.000 li de largo, es decir, 10.000 kilómetros.

Texto: chfo06

### La terminación dialectal **er**

Puede constituir por sí sola una sílaba, pero también puede unirse a una vocal o grupo vocálico precedente para formar una vocal retrofleja. Al escribirlo en *pinyin* o pronunciarlo, solemos añadir una “r” detrás de la vocal o grupo de vocales como *nǎr* o *nàr*.

Por ejemplo: [el vino de España]

那西班牙的那红酒不是特别有名**儿**吗？你没在这**儿**多喝点**儿**？还是你根本就不喜欢红酒，你觉得它度数太低，没意思？

nà xī bān yá de nà hóng jiǔ bú shì tè bié yǒu **míng**  
ma? nǐ méi zài **zhèr** duō hē **diǎnr**? hái shì nǐ gēn běn  
jiù bù xǐ huān hóng jiǔ, nǐ jué de tā dù shu tài dī,  
méi yì si?

¿No es el vino de España muy famoso? ¿No has bebido más estando aquí? ¿O es que no te gusta nada el vino? ¿Te parece que tiene pocos grados, y que no es interesante?

Texto: chin01

## 7.3 El uso de corpus en la enseñanza de comprensión auditiva

Las muestras orales nos facilitan desarrollar las destrezas de comprensión auditiva de los alumnos y diseñar actividades en clase.

Tomamos un texto corto como ejemplo:

- 1 0-4.709 AZY: 麻烦问一下 / 橙子多少钱一斤 ?
- 2 4.709-6.904 YAP: 橙子两块五一斤 ///
- 3 6.904-9.227 AZY: 噢 // 旁边的呢 ?
- 4 9.227-13.521 YAP: 旁边这个是澳柑 // 澳柑是八块八一斤 ///
- 5 13.521-15.477 AZY: 怎么会这么贵呀 ?
- 6 15.477-18.188 YAP: 它这个澳柑是进口的 ///
- 7 18.188-21.421 YAP: 你看 / 它这个包装也比较精致 ///
- 8 21.421-24.584 YAP: 味道也应该是好一些 [///] 好的很多呢 !
- 9 24.584-26.995 AZY: 噢 / 算了 / 还是太贵了 ///
- 10 26.995-28.729 AZY: 我 [/] 我买二斤橙子吧 !
- 11 28.729-29.48 YAP: 好嘞 !
- 12 29.48-32.738 YAP: 哎 / 这个香蕉今天是打特价 ///
- 13 32.738-36.737 YAP: 平时是两块五 / 今天两块二 / 买一点吗 ?
- 14 36.737-37.839 AZY: 是吗 ?
- 15 37.839-38.886 AZY: 拿来我看一下 ///
- 16 38.886-39.365 YAP: 呐 {%com: give the bananas to him} ///
- 17 39.365-42.304 AZY: 噢 {%act: assent} ///
- 18 42.304-44.208 YAP: 挺好的 / 质量 ///
- 19 44.208-46.309 AZY: 那给我约二斤吧 ///
- 20 46.309-49.832 YAP: 噢 // 这个苹果 [/] 今天的苹果也是不错的  
///
- 21 49.832-52.123 YAP: 这个是正宗的红富士 ///
- 22 52.123-53.115 AZY: 噢 ///
- 23 53.115-53.821 YAP: 也要吗 ?
- 24 53.821-54.973 AZY: 算了 / 不要了 ///
- 25 54.973-57.605 AZY: 我一个人吃 / 吃不了太多的 ///
- 26 57.605-61.23 YAP: 噢 / 小伙子 // 我看你有点上火 ///
- 27 61.23-62.909 YAP: 是不是应该买点梨 ?
- 28 62.909-69.214 YAP: 买梨子回去用冰糖跟梨子炖水 // 这是下火的  
///
- 29 69.214-70.216 AZY: 是吗 ?
- 30 70.216-72.919 YAP: 恩 {%act: assent} / 你可以煮一煮 / 可以  
下火 ///
- 31 72.919-74.476 AZY: 那给我约三个吧 !

32 74.476-75.644 YAP: 好 ///

33 75.644-77.438 AZY: 不要放太 [ / ] 不要约太多 ///

34 77.438-78.074 YAP: 好 ///

35 78.074-78.958 AZY: 回头放坏了 ///

36 78.958-79.841 YAP: 嗯 ///

37 79.841-84.016 AZY: 另外问一下子 // 水 {%pho: [suĩ]} 果是在这里结帐还是在款台结帐 ?

38 84.016-87.367 YAP: 水果都在收银台结帐 // 统一 ///

39 87.367-90.934 AZY: 噢 / 再麻烦问一下子 / &mm 牙膏在哪里呀 ?

40 90.934-94.341 YAP: 牙膏在二楼的日用品区 ///

41 94.341-96.854 AZY: 好 / 那谢谢 / 再见 !

42 96.854-98.63 YAP: 嗯 / 下次再来 !

43 98.63-100.04 AZY: 好 / 再见 !

Figura 7.11 Ejemplo de supermercado de C-ORAL-CHINA

Con este diálogo podemos diseñar una serie de actividades para practicar la comprensión auditiva de los alumnos.

1. Los alumnos escuchan la grabación y contestan a las preguntas acerca de las informaciones del diálogo, por ejemplo, ¿dónde están?, ¿qué relaciones tienen los dos participantes?, ¿qué compró al final el chico? , etc.
2. El paso siguiente puede ser dejarles un ejercicio de rellenar el blanco para que puedan volver a escucharla y captar los detalles de esta conversación.
3. Después de entender el contenido, los alumnos pueden leer la transcripción y estudiar relación entre formas lingüísticas y funciones lingüísticas. El profesor puede dirigirles a observar qué elementos se utilizan en este diálogo para hacer comparación, expresar acuerdo, vacilación y rechazo, despedirse, etc.

4. El profesor puede dirigir a los alumnos a detectar la diferencia del uso de palabras entre el español y el chino. En este caso, podemos hacer comparación entre “斤” (medio kilo) y kilo. En China, la gente suele medir el peso utilizando “斤”, que corresponde a medio kilo. También puede explicarles la diferencia entre “两” y “二”, como hemos planteado en el apartado 7.2.

5. Los alumnos pueden elaborar un listado de palabras relacionadas con fruta y peso en chino. Luego, se puede organizar actividades de interacción para practicarlas. Por ejemplo, los alumnos pueden intentar imitar este diálogo y diseñar otro parecido entre vendedor y comprador. Uno trata de usar las palabras estudiadas para aconsejar al otro que compre lo que quiera vender. Y el otro puede practicar cómo preguntar el precio, hacer regateo y la decisión final. Después de dominar las palabras de fruta, el profesor puede darles otro listado de palabras de otros sectores para que sigan practicando la destreza de la compraventa.

#### ***7.4 El uso de corpus en la enseñanza de fonemas***

En la transcripción de C-ORAL-CHINA, hemos elaborado un subcorpus y hecho la transcripción de *pinyin*.

La tabla 7.12 muestra los datos básicos de este subcorpus. En total, son cinco textos que duran más o menos una hora. Tiene 13.267 caracteres.

<b>Archivo</b>	<b>Tipo</b>	<b>Caracteres</b>	<b>segundos</b>
chfo01	Formal	4349	1334
chfo03	Formal	2096	611
chin13	Informal	4594	980
chme12	Medios de comunicación	304	96
chme20	Medios de comunicación	1924	540
Total		13267	3561

Tabla 7.12 Datos básicos del subcorpus de *pinyin*

Según el subcorpus, hemos calculado la frecuencia de los símbolos fonográficos *pinyin*. Los profesores pueden decidir el orden de la enseñanza de letras de *pinyin* según el listado. Los alumnos pueden buscar según *pinyin*, escucharlas para mejorar la pronunciación, estudiar la composición de la sílaba en *pinyin* y hacer la comparación con el español.

Empleando *pinyin*, podemos observar la configuración fonológica del chino. Hay 21 consonantes iniciales de sílaba. La tabla 7.13 nos muestra la frecuencia y el porcentaje que ocupa cada uno en el subcorpus. En contraste, en español, hay 19 fonemas de consonantes iniciales (b, p, m, f, d, t, n, l, g, k, j, ch, ll, ñ, y, r, rr, s, z). Es fácil notar la diferencia entre los dos idiomas.



<b>Consonante inicial</b>	<b>Número</b>	<b>Porcentaje</b>
n	1590	12.13%
d	1442	11.00%
g	1327	10.13%
sh	1098	8.38%
h	976	7.45%
s	876	6.68%
j	841	6.42%
zh	764	5.83%
x	763	5.82%
z	629	4.80%
b	558	4.26%
l	554	4.23%
m	507	3.87%
t	497	3.79%
q	326	2.49%
k	298	2.27%
r	290	2.21%
f	241	1.84%
ch	211	1.61%
p	119	0.91%
c	54	0.41%
En total	13961	

Tabla 7.13 Frecuencia de consonante inicial

<b>Vocal</b>	<b>Número de tono ligero</b>	<b>Número de primer tono</b>	<b>Número de segundo tono</b>	<b>Número de tercer tono</b>	<b>Número de cuarto tono</b>
a	290	821	686	641	1281
o	670	409	190	703	517
e	1478	279	690	434	671
i	2640	421	621	607	870
u	1877	128	239	208	619
ü			0	5	4
ai	25	49	247	60	211
ei	0	32	106	46	108
ao	31	68	20	207	304
ou	20	81	26	321	248
an	79	357	371	305	480
en	141	137	263	182	82
ang	54	138	90	148	178
eng	0	84	99	22	33
ong	3	145	78	97	118
ia	48	217	75	211	332
ie	2	99	51	39	18
iao	17	28	5	80	117
iu	1	11	32	3	241
ian	23	102	68	131	169

in	12	169	113	53	108
iang	4	17	10	61	43
ing	12	74	83	24	75
iong	0	3	4	0	0
ua	12	65	45	37	84
uo	5	176	78	69	140
uai	0	1	3	0	9
uan	3	62	44	37	44
uang	2	5	3	1	14
üe	0	0	166	4	40
ün	0	9	15	7	15

Tabla 7.14 Frecuencia de vocal según el tono

<b>Vocal</b>	<b>Número</b>	<b>Porcentaje</b>
i	4584	26.48%
e	2192	12.66%
u	1958	11.31%
o	1384	8.00%
a	1071	6.19%
ai	579	3.34%
en	567	3.28%
uo	468	2.70%
ou	408	2.36%
ao	383	2.21%
ian	358	2.07%
ang	308	1.78%
an	301	1.74%
ei	292	1.69%
iou(iu)	288	1.66%
ing	268	1.55%
iao	247	1.43%
eng	238	1.37%
üe	210	1.21%
ie	209	1.21%
in	187	1.08%
ong	173	1.00%
uan	165	0.95%

ia	143	0.83%
iang	135	0.78%
ün	46	0.27%
ua	40	0.23%
uang	25	0.14%
uai	13	0.08%
ü	9	0.05%
iong	7	0.04%

En total 17256

Tabla 7.15 Frecuencia y porcentaje total de vocal

Como es sabido, el chino es una lengua tonal. Los tonos se colocan encima de vocal. Esto se distingue del acento de español. La tabla 7.14 y 7.15 nos presentan distintas estructuras de vocal, el porcentaje que ocupan y la frecuencia de cinco tipos de tonos. En comparación con el español, la vocal de chino es más complicada. Se pueden detectar pronunciaciones que no existen en español y encontrar ejemplos reales en el corpus para practicarlas. Además, tanto el chino como el español tienen vocales compuestas (diptongo y triptongo). Sin embargo, las estructuras son limitadas y hay que dominarlas. En chino, hay muchas sílabas que tienen vocales compuestas. Si los alumnos no pueden pronunciarlas correctamente, afectará al entendimiento y al intercambio.

Aparte, cabe señalar que el chino no tiene el fenómeno de enlace fónico, lo cual debe llamar la atención a los hispanohablantes que están acostumbrados a ello.

## 8. Conclusión

En los últimos años, la Lingüística de Corpus se ha desarrollado rápidamente, abriendo un nuevo camino en la investigación lingüística. Cada vez hay más lingüistas que se fijan tanto en su utilidad como en su potencialidad. El progreso permanente de la tecnología informática y la popularización de ordenadores personales han hecho que todo el mundo tenga posibilidad de acceder a los valiosos recursos de los corpus. Su introducción en el marco de la renovación de metodologías de la enseñanza de las lenguas ha adquirido mucha importancia.

En este trabajo, hemos creado el corpus de habla espontánea de chino *putonghua* C-ORAL-CHINA. Es un proceso costoso. Hemos decidido crearlo por diferentes razones. Comprende materiales didácticos muy útiles y puede reflejar los rasgos caracterizadores de la lengua oral o de registros que no aparecen en otros recursos didácticos. Nuestra investigación ha sido proceso lleno de creatividad, en el que pudimos controlar todo procedimiento desde el diseño hasta las propuestas de sus aplicaciones para la enseñanza. En comparación con los corpus generales, los corpus especializados pueden ser más factibles para nuestro objetivo didáctico. Podemos hacer la selección de temas nosotros mismos. Estamos más familiarizados con el contenido del corpus, lo cual sin duda facilita mucho el procesamiento en el futuro y el diseño de actividades educativas. Las aplicaciones de corpus en la enseñanza son muy amplias; sin embargo, no están todavía muy desarrolladas. La investigación en este aspecto es necesaria y urgente. Cómo usarlo de manera razonable y eficiente es el tema de estudio y discusión con el que nos enfrentamos.

C-ORAL-CHINA cumple con todas las características que debe poseer un corpus de habla espontánea actual. En primer lugar, su formato es electrónico, permitiendo el almacenamiento y la manipulación de datos. Tiene una estructura propia de clasificación

de muestras. Se divide en tres subcorpus: formal en contexto público, medios de comunicación e informal. Además abarca una amplia variedad de contextos *situacionales* (monólogos y diálogos), de *registro* (formal e informal) o de *contenido* (conferencias sobre la cultura china, entrevistas en medios de comunicación, etc.). Por último, su presentación en una página web facilita su disponibilidad para aquellos que estén interesados en el aprendizaje del chino *putonghua*.

En total, C-ORAL-CHINA consta más de 140.000 caracteres. Siguiendo la metodología de LLI-UAM, se caracteriza por la espontaneidad.

El recurso se presenta en diferentes formatos:

1. Versión sonora;
2. Versión de texto que incluye metadatos acerca de los participantes, transcripción ortográfica y la transcripción de *pinyin* (de un subcorpus extraído de C-ORAL-CHINA);
3. Versión etiquetada en XML;
4. Alineamiento entre sonido original y transcripción ortográfica.

Además, planteamos nuestras propias propuestas didácticas sobre el uso de corpus en la enseñanza de *pinyin*, léxico, gramática y comprensión auditiva.

Al final, en los **Apéndices**, aportamos los resultados cuantitativos obtenidos en este trabajo de investigación. Resultan útiles para realizar estudios lingüísticos, comprobar hipótesis o hacer comparaciones con otros corpus u otras variedades lingüísticas.

Para su consulta, los profesores y alumnos de chino *putonghua* pueden visitar la página web<sup>12</sup>:

<http://drusila.llf.uam.es/lab/?q=c-oral-china>

Comprende los siguientes apartados:

- Transcripción: en esta sección se facilitan los metadatos y el texto total de la transcripción alineado con su correspondiente sonido.
- Búsqueda: la aplicación *Concordancias* permite al usuario consultar cualquier palabra o estructura en sus contextos de aparición de este corpus. De esta forma, se pueden localizar los enunciados de este corpus que tienen dicha palabra junto con su sonido y una breve indicación para que los usuarios puedan consultar el texto original fácilmente.

En cuanto al trabajo futuro, las líneas de investigación quedan abiertas hacia cuatro tareas no menos importantes:

1. La transcripción en *pinyin* de todo el corpus.
2. La ampliación del tamaño del corpus e inclusión de una mayor variedad de situaciones comunicativas.
3. Profundización en la investigación del uso de corpus en la enseñanza, relacionando la teoría con la práctica.
4. Realización de una comparación exhaustiva en los niveles fonológico y léxico entre el español oral y el chino oral.



## Bibliografía

Adolphs S., Knight D. (2010): “Building a spoken corpus: what are the basics?” En: McCarthy M. y O’Keeffe A. (Eds.) *The Routledge Handbook of Corpus Linguistics*. Nueva York, Routledge.

Aijmer, K., Altenberg, B. (ed.) (1991): *English Corpus Linguistics. Studies in honor of Jan Svartvik*. London, Longman.

Alderson C. (1996): “Do corpora have a role in language assessment?” En: Thomas J. y Short M. (ed.) *Using Corpora for Language Research*. London, Longman.

Braun, S. (2005): “From Pedagogically Relevant Corpora to Authentic Language Learning Contents”. *ReCALL* 17(1), p. 47–64

Briz, A. (2002): *Corpus de Conversaciones Coloquiales*. Madrid, Arco-Libros.

Campillos L., Gozalo P., Guirao J., Moreno A. (2010): *Español Oral en Contexto*. Madrid, UAM Ediciones.

Campillos L., Gozalo P., Guirao J., Moreno A.: “Exploiting a Spoken Corpus in Language Teaching/Learning: An Advanced Web-based Tool”. [en línea]: < <http://www.mavir.net/publications> > [consulta: 28 de febrero de 2011]

Cao Wei (2003): “现代汉语口语词和书面语词的差异初探”. *语言教学与研究* 2003 (6). Beijing, Editorial de Enseñanza e investigación de lenguas.

Carter R., McCarthy M. (1995): “Grammar and the spoken language”. En: *Applied Linguistics*, XVI, n. 2, p.155

Chen Heqin (1928): *语体文应用字汇*. Shanghai, the Commercial Press, p.1-10

Cheng W. (2010): “What can a corpus tell us about language teaching”. En: McCarthy M. y O’Keeffe A. (eds.) *The Routledge Handbook of Corpus Linguistics*. Nueva York, Routledge.

Cheng W., Warren M (2005): “A Corpus-driven Study of Disagreement in Hong Kong Business Discourse”. En: M. Gotti y F. Bargiela (ed.) *Asian Business Discourse(s)*. Frankfurt am main, Peter Lang, p. 241–270

“Chinese, mandarin”. En: *Ethnologue Languages of the World* [en línea]:< [http://www.ethnologue.com/show\\_language.asp?code=cmn](http://www.ethnologue.com/show_language.asp?code=cmn)> [consulta: 15 de febrero de 2011]

Comrie B. (1990): *The World’s Major Languages*. New York, Oxford.

Cresti E., Moneglia M. (2005): *C-ORAL-ROM Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia, John Benjamins Publishing Company.

Ding Xinshan (2005): “语料库语言学的发展及研究现状”. En: Yi Mianzhu y Nan Zhenxing (ed.) *Computational Linguistics*. Shang Hai, Shanghai Foreign Language Education Press.

Feng Zhiwei (2008): *Advances in Corpus linguistics*. Beijing, WPC.

Fisac T. (2008): “Algunas propuestas metodológicas para la enseñanza de la lengua china: planteamientos iniciales para una ulterior discusión.” En: *La enseñanza de la lengua china como lengua extranjera en España*. Valladolid, Ayuntamiento de Valladolid, p.40.

Fligelstone, S. (1993): “Some Reflections on the Question of

Teaching, from a Corpus Linguistics Perspective”. *ICAME Journal* 17, p. 97–110

Flowerdew, L. (2002): “Corpus-based Analyses in EAP”. En: C. Candlin (ed.) *Academic Discourse*. London, Longman.

Font Rotchés, Dolors (2007): “Patrones Melódicos del Catalán en Habla Espontánea”. En: *Actas del III Congreso de Fonética Experimental*. Santiago de Compostela, Xunta de Galicia.

Gao Jianping, Gao Xiaomei, Cheng Shuming (1980): *汉语发展史*. Harbin, Harbin Engineering University Press.

Granger S., Dagneaux E., Meunier F. (2002) : *International Corpus of Learner English*. Louvain, UCL Presses Universitaires de Louvain.

Greenbaum S., Svartvik J.: “The London-Lund Corpus of Spoken English” [en línea]: < <http://icame.uib.no/london-lund/>> [consulta: 13 de abril de 2011]

HarperCollins: [en línea]:  
< <http://www.collins.co.uk/Corpus/CorpusSearch.aspx> > [consulta: 3 de abril de 2011]

Hu Jicheng: “《汉语拼音方案》述评”. [en línea]:  
<<http://library.crtvu.edu.cn/wb/ReadNews.asp?NewsID=3871>>  
[consulta: 5 de febrero de 2011]

Hunston S. (2002): *Corpora in Applied Linguistics*. Cambridge, Cambridge University Press

Hunston S. (2006): *Corpora in Applied Linguistics*. Beijing, WPC press

Johns, T. & P. King (eds.) (1991): *Classroom Concordancing*. Birmingham, University of Birmingham.

International Corpus of English: [en línea]: <  
<http://ice-corpora.net/ice/index.htm> > [consulta: 13 de abril de 2011]

Laboratorio de Lingüística Informática [en línea]: <  
<http://www.lllf.uam.es/ESP/>> [consulta: 16 de enero de 2011]

Leech G. (1997): “Teaching and language corpora: a convergence”.  
En: *Teaching and Language Corpora*. London, Longman.

Li Aijun, Zheng Fang, Byrne W., Fung P., Kamm T., Lui Yi, Song Zhanjiang, Ruhi U., Venkataramani V., Chen Xiaoxia (2000): “Cass: A Phonetically Transcribed Corpus of Mandarin Spontaneous Speech.” Beijing, International Conference on Spoken Language Processing.

Li Xingjian (2004): *现代汉语规范词典*. Beijing, Editorial de Enseñanza e Investigación de Lenguas Extranjeras.

Liu Yabin, Li Aijun: “朗读语料与自然口语的差异分析” [en línea]:  
< [http://ling.cass.cn/yuyin/report/files/2002\\_9.pdf](http://ling.cass.cn/yuyin/report/files/2002_9.pdf) > [consulta: 3 de abril de 2011]

Mackin, E. (2003): “Romancing the Rosetta Stone.” [en línea]:  
<[http://www.eurekalert.org/pub\\_releases/2003-07/uosc-rtr072403.php](http://www.eurekalert.org/pub_releases/2003-07/uosc-rtr072403.php)> [consulta: 3 de enero de 2011]

McCarthy M. y O’Keeffe A. (eds.) (2010): *The Routledge Handbook of Corpus Linguistics*. Nueva York, Routledge.

McCarthy M., O’Keeffe A. (2010): “Historical perspective---What are corpora and how have they evolved?” En: McCarthy M. y O’Keeffe A. (eds.) *The Routledge Handbook of Corpus Linguistics*. Nueva York, Routledge.

McEnergy T.(2009): “Where are we going?”. En: *Teaching and*

*Learning by Doing Corpus Analysis*. Beijing, WPCBJ, p.4

McEnery T., Xiao R., Tono Y.(2006): *Corpus-Based Language Studies*. Nueva York, Routledge.

McEnery T., Xiao R.: “Quantifying Constructions in English and Chinese A Corpus-Based Contrastive Study”. [en línea]: <  
[http://www.corpus.bham.ac.uk/corplingproceedings07/paper/19\\_Paper.pdf](http://www.corpus.bham.ac.uk/corplingproceedings07/paper/19_Paper.pdf)> [consulta: 1 de enero de 2011]

Ministerio de Asuntos Exteriores de China: “中国同西班牙的关系”. [en línea]:

[http://www.mfa.gov.cn/chn/gxh/cgb/zcgmzysx/oz/1206\\_41/1206x1/t7362.htm](http://www.mfa.gov.cn/chn/gxh/cgb/zcgmzysx/oz/1206_41/1206x1/t7362.htm) [consulta: 25 de noviembre de 2010]

Moreno A. (1998): *Lingüística Computacional*. Madrid, Editorial Síntesis.

Moreno A., Urresti J. (2005): “El Proyecto C-ORAL-ROM y su aplicación a la enseñanza del español”. *Oralia. Análisis del discurso oral*. Madrid, Arco Libros, p. 81-104

Norman J. (2008): *Chinese*. Beijing, World Publishing Corporation.

Oficina Nacional de Propaganda Internacional de chino: “汉语中词类与句法成分之间不是简单的对应关系”. [en línea]: <  
[http://www.xmuoec.com/gb/teacher/yufa/02/02\\_03\\_2.htm](http://www.xmuoec.com/gb/teacher/yufa/02/02_03_2.htm) >  
[consulta: 3 de enero de 2011]

Patrice Pavis (1998): *Diccionario de teatro: dramaturgia, estética, semiología*. Barcelona, Paidós.

Santalla del Río, M. P. (2005): “La Elaboración de corpus lingüístico”. En: M. Cal Varela (coord.) *Nuevas tecnologías en lingüística, traducción y enseñanza de lenguas*. Santiago de Compostela, Universidade de Santiago de Compostela.

Renouf, A. (1997): “Teaching Corpus Linguistics to Teachers of English”. En: Wichmann A., Fligelstone S., McEnery T. y Knowles G. (eds) *Teaching and Language Corpora*. London, Longman, p. 255–66.

Resourced Pool for Teaching and Promoting Chinese as a Foreign Language: “现代汉语词汇的特点”. [en línea]: <<http://www.hy123.org/zhishi/word/290.html>> [consulta: 11 de enero de 2011]

Römer U. (2005): *Progressives, Patterns, Pedagogy. A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam, John Benjamins, p. 261-263

Shi Youwei(2000): *汉语外来词*. Beijing, The Commercial Press, p. 1-20

Sinclair, J. McH. (2004): “Introduction”. En: J. M. Sinclair (ed.) *How to Use Corpora in Language Teaching*. Amsterdam y Philadelphia: John Benjamins.

Sun Yizhen(1999): *新汉西词典*. Beijing, The Commercial Press, p.232, 887

Tian Xiaolin: “现代汉语词汇的特点”. [en línea]: <<http://www.huayuqiao.org/articles/tianxiaolin/txl12.htm>> [consulta: 11 de enero de 2011]

Transana: [en línea]: <[www.transana.org](http://www.transana.org)> [consulta: 25 de noviembre de 2010]

Villayandre M.: “Lingüística Computacional II”. En: *Curso monográfico sobre Lingüística de corpus* [en línea]: <<http://www3.unileon.es/dp/dfh/Milka/LCII/Corpus4.pdf>> [consulta: 3 de abril de 2011]

VV.AA. Departamento de Chino del Comité de Lenguaje Hablado y Escrito: “现代汉语常用字表”. [en línea]: <<http://202.205.177.9/edoas/website18/40/info16840.htm>> [consulta: 3 de febrero de 2011]

VV.AA. Oficina del Consejo Internacional del Idioma Chino (2006): *Conocimientos Comunes de la Cultura China*. Beijing, Editorial Sinolingua.

Wang Li (1980): *汉语诗稿*. Beijing, Zhonghua Shuju.

Wen Songhui: “雅言、正音、官话 古往今来的“普通话””. Diario de Liberación. [en línea]: <<http://book.people.com.cn/GB/69398/13474884.html>> [consulta: 5 de febrero de 2011]

Wlater O.(1987): *Oralidad y escritura. Tecnologías de la palabra*. México, Fondo de Cultura económica.

Yin Binyong: “Xizi Qiji” [en línea]: <<http://www.chinabaike.com/article/sort0525/beijingbaike/2007/20070722145169.html>> [consulta: 5 de febrero de 2011]

Yuan Jiahua (2001): *汉语方言概要*. Beijing, Editorial de Chino, p. 22

Zhang Yanqun, Gao Qiliang (2008): *现代汉语语法通论*. Zheng Zhou, Editorial de Pueblo de He Nan, p. 10-14

Zhao Shiyu (1999): *Comparación Bilingüe entre el Chino y el Español*. Beijing, Editorial de Enseñanza e Investigación de Lenguas Extranjeras, p.22

“《当代北京口语语料库》说明”: [en línea]: <[http://www.blcu.edu.cn/yys/6\\_beijing/yuliao/shuoming.htm](http://www.blcu.edu.cn/yys/6_beijing/yuliao/shuoming.htm)> [consulta: 17 de abril de 2011]

## *Bibliografía*

“六书”: [en línea]:

<<http://zh.wikipedia.org/zh-cn/%E5%85%AD%E6%9B%B8>>

[consulta: 2 de febrero de 2011]



## Apéndice A

### *Ejemplos de transcripción de C-ORAL-CHINA*

#### Transcripción ortográfica

##### 1. Ejemplo del subcorpus Formal

[  
@Title: aprendizajechino  
@File: chfo01  
@Participants: CRJ, (woman, B, 3, profesor, participant)  
@Date: 28/10/2009  
@Place: Madrid  
@Situation: in a classroom, not hidden  
@Topic: Textbooks, prep, review, customs, grammar  
@Source: C\_ORAL\_CHINA  
@Class: formal, public, class  
@Length: 22' 14"  
@Characters: 4444  
@Acoustic\_quality: A  
@Transcriber: Dong Yang  
@Revisor: AZY  
@Comments:  
]

\*CRJ: 同学们好 ! α<2785> 今天是我们上课的第一天 ///α<5232> 我  
先来介绍一下我们这门课 // 怎么 hhh {%com: close the door} 来上  
///α<9907> 这一门课我们学习的是 // 新实用汉语课本的第二册  
///α<15606> 这个学期 / 我们学习 / 第十五课 / hhh {%com: open the  
book} 到第二十课 ///α<21853> 一共我们学习六课书 ///α<24330> &mm  
每一课书基本上是用六个小时 ///α<28957> 前 [/ 前四个小时或者五个小  
时 // 我们学习 / 生词 / 语法 / 课文和一部分的练习 ///α<38243> 最后

一个小时或者两个小时 // 我们做口头的 / 表达练习 ///α<43516> 我来介绍一下这一 [//] 我们怎么来上 / 这个课 ///α<51368> &mm 首先呢 // 我们每一课中都是要先学习生词 ///α<56011> 生词有一些词呢 / 大家已经学习过了 // 已经认识了 ///α<60390> 这个我们就会讲得比较 / 少一些 ///α<62881> 有一些重点的词呢 // 我们会 / 举一些例子 // 然后让大家多说一些句子 ///α<68647> &mm 然后呢 / 是 // &mm 讲 [//] 讲解课文 ///α<73839> 课文呢 // 我们主要要多读 ///α<76091> 上课的时候我们没有很多的时间来读课文 ///α<79255> 所以需要大家下课以后 // 反复地读 [//] 读课文 ///α<84221> 最好是能达到背诵的程度 ///α<86726> 那么上课呢 // 我们就读了课文以后 // 理解了课文的意思 // 回答一些 / 问题 // 根据课文回答一些问题 ///α<96896> 然后呢 // 我们的 [//] 其实的 / 讲课最重要的一个重点是 // 语法的部分 ///α<104191> 这个语法呢 / 每一课都有几个语言点 ///α<107555> 这几个语言点呢 / 我们会 // 逐一来讲解 ///α<112445> 然后呢 / 做很多的练习 ///α<114560> hhh {%act: clearing of throat} 上课的时候 // 我们希望大家就是多说 // &mm 就是提高开口率 ///α<120694> &mm 然后呢 // 是课后的练习 ///α<125474> 因为我们这个书的课后练习比较多 // 所以我们不可能在上课的时间 // 每一个练习都做 ///α<132101> 我会列出一些重点的练习 // 大家上课做 // 或者是有一些留作作业 ///α<139388> 每一课呢 // 我们会有一个话题 ///α<142241> 这个话题讨论呢 / 是需要大家 // 提前准备 ///α<146394> 比如说我布置给你一个作业 ///α<148467> 你 // 两个同学或者一个同学 // 这个我们根据每一课不同的情况 // 来 / 决定 ///α<157654> 两个同学一组 // 或者是一个同学独立完成 ///α<161240> 来就这个话题写一个 / 小的文章 // 或者一个对话 ///α<166511> 然后上课的时候呢 // 你需要到前边来表演 ///α<170945> 这个主要是为了练习大家的 / 口头表达能力 ///α<176781> &mm 关于 [//] 这是关于我们上课的 / 步骤 ///α<184402> 那么关于考试 // 是这样的 ///α<188107> 我们的考试呢 / 分成三个部分 ///α<190632> 第一个部分呢 / 是笔试 ///α<193063> 这个 // &mm 占 70% ///α<198142> 第二部分是口试 // 占 20% ///α<202363> 第三部分就是平时的作业 // 这个占 10% ///α<206427> 所以 // 如果 / &mm 你平时的作业不交的话 // 那么 // 你最后的成绩可能就不会特别好 ///α<217583> 而我觉得学习中文呢 // 因为大家的 [//] 现在在西班牙的这个语言环境里 / 可能没有太多的 / 机会来说汉语 ///α<229573> 但是 / 我觉得我们可以通过网络呢 / 来看一些中文的电影 // &mm 或者是 / 我们这个课文呢 // 也有 / 这个 / CD ///α<240516> 大家可以一边 [//] 一边走路啊 // 或者一边做别的事情的时候 // 可以一边听这个课文的 CD // 或者是听一些中文歌曲

///α<251325> 或者有时间的时候 // 看一点儿中文电影 ///α<253843> 这样  
 对于提高听力 // 是很有帮助的 ///α<259771> 再有一个呢 / 我希望大家  
 就是上课的时候 // 尽量都说汉语 ///α<266483> 包括课间 [/] &mm 课间  
 休息的时候 // 如果我们学过的句子 // 学 [///] 可以说的句子 // 简单的  
 句子 // 尽量用汉语来表达 ///α<277988> 还有呢 // 就是需要预习和复习  
 ///α<281371> 这个是 / 特别需要强调的 ///α<285506> 就是 [///] 因为我  
 们的时间 // 每一课 [/] 每一个星期只有四节 [/] 四个小时的汉语课  
 ///α<292388> 本来时间就很短 // 如果你不复习不预习的话 // 很容易就  
 忘掉了 ///α<297400> &mm 这是 =α<301202> hhh {%com: open the book}  
 我再讲一下这本书 [/] 它的这本书的 / 这个 // 主要的安排的 / 想法 //  
 就是它的主要的想法 ///α<315004> 就是这一课 [///] 这一本书呢 // 从 [/]  
 这个 [/] 从这个 &mm 作者来编写这本书的时候呢 // 他考虑的 / 有这  
 么几个方面 ///α<330712> 第一个呢 / 就是说你在语法 / 教学的阶段 //  
 关于重点的句型 // 还有重点的词语 // 这个我们就是觉得 // 成年人的学  
 习第二种 [///] 成年人来学习第二种语言 // &mm 不完 [///] 不可能完全  
 是一种自然的习得 ///α<350019> 只是盲目地模仿或者死记硬背 // 可能  
 达不到很好的 / 学习效果 ///α<358562> 所以呢 / 必须要对 // 课文当中  
 出现的词语和句型加以控制 // 来通过一定的规则 // &mm 组词组句的一  
 些规则 // 来达到掌握语言的目的 ///α<373217> 特别是西班牙语和汉语  
 呢 // 有很不同的特点 ///α<377288> 有 [/] 有很明显的不同的特点  
 ///α<380083> 所以呢 / 就需要 / 这个学习的时候呢 // 有一定的系统性  
 // 来引导学生 [/] &mm 引导学习者呢 // 有正确的总结词语组句的一个  
 规则 // 来进行交际 ///α<394817> &mm 这个是在 [///] 是 / 应该比较强  
 调的一点 ///α<401428> 就是说 // 学习的 [///] 学习汉语的时候 // 要 [/]  
 需要按照一定的规则 // 汉语的规则 [/] 语言规则来 [/] 来学习  
 ///α<409922> 不能是随便地把一些词语组在 [/] 组在一起 ///α<415111>  
 这样 / 学出来的语言呢就是不准确的或者 [/] 或者说不扎实的  
 ///α<420925> 所以 / 我们这本书的每一课的这个重点呢 // 都是 [///] 每  
 一课课文里出现的重点都是 // 体现在语法的部分 ///α<431513> &mm 这  
 些语法 // 在课文当中呢 // 就会有多次的重复 // 重现 ///α<439318> 所  
 以呢 // 就是以前学过的重点句型和语法 // 在这一课中 // 可能也会多次  
 的重复 ///α<446173> 所以 // 在 [///] 所以我强调大家 &学 // 反复地 /  
 多读课文 // 多背课文 // 也就是在这一方面 ///α<454585> 如果 / 你能够  
 把课文很熟练地说出来 // 那么 / 其中的这个语言点 // 你也就掌握了  
 ///α<463108> &mm 好 // 还有呢 // 就是这个练习和 [///] 练习的部分 [/]  
 &mm 练习的部分呢 // 我们需要 // &mm 有重点地进行 [/] &mm 进行

这个选择 [482675](#) 主要是围绕着这一课重要的词汇和语法 [489163](#) &mm 还有呢 // 就是这个生词 [492723](#) 生词的部分 // 每一 [499608](#) 这一 [506830](#) 每一课的生词呢 / 比较多 // 大概有 40 个左右 [506830](#) 所以呢 // 要 / 注意 // 选择重点的词语来掌握 [506830](#) 如果一些不是很常用的 // 或者是书面语 [ ] 书面语色彩特别强的词语 // 可以暂时不用 [ ] 不用掌握 [521169](#) &mm 还有汉字 / 对 // 这个对 &mm 西班牙的同学来说呢 / 也是一个难点 [528691](#) 就是汉字是 [530962](#) 汉字应该怎么学 ? [532643](#) 这个我们在这本 [ ] 这本 [ ] 这本书呢 / 它每一课的那个练习册上 [ ] 学生的练习册上 // 它都有汉字的练习 [541724](#) 你可以模仿着这个 / 书中列出的 // 这些汉字 / 来一笔一画地学 [ ] 学习 [549548](#) 如果汉字不能 &mm 这个快速的 [ ] 不能这个掌 [ ] 不能掌握一定的汉字的话 // 你在今后学习汉语的过程中 // 就会遇到很多的困难 [563875](#) 在包括阅读上 // 还有这个 &mm 理解方面 // 会遇到很多的困难 [568879](#) 所以不能说因为汉字 / 对大家来说比较难 // 就放弃了 [573826](#) 必须坚持 / 写汉字和认读汉字 [577820](#) &mm 口语听力的水平呢 / 这个 // 刚才 &mm 我说 // 就是听力呢 // 我说 / 可以大家用课外的时间 // 适当地看一些 // 中国的电影 // 或者是听中国歌曲 // 或者听我们课文的这个 &mm CD [600589](#) 还有呢 // 就是口语的方面呢 // 你可以 [604296](#) 因为现在我知道有很多中国留学生在这儿 // 所以大家如果有 [ ] 有可能的话 // 可以找一些中国留学生 // 认识一些中国留学生 [613803](#) 平时呢 // 跟他们聊天儿 [616340](#) 你 [ ] 你们 [ ] 你可以这个说西班牙语 // 然后他说汉语 [622224](#) 这种自然的聊天儿的过程呢 // 也是很好的一个学习方式 [627962](#) 还有我觉得背课文也是一个好办法 [635487](#) 就是说背课文呢 / 也是学习语言的一个好办法 [638817](#) 就是你通过背课文呢 // 这些语言点 // 可以自然地记在你的脑子里 // 而且你会 &培 [ ] 那个培养出正确的一个语感 [647569](#) &mm 这个能够帮助你自然流利地表达 // 而不需要用一个语言规则 // 这样去套它 [657858](#) 这个 [ ] &mm 这个是我们的教材 // 当然还有一些其他的参考教材 [663782](#) &mm 我 [ ] 我可以告诉大家 // 但是我不知道你们可不可以在西班牙找到 // 可能不是很容易 [670725](#) 我已 [ ] 我已经从汉办定了一些教材 [675500](#) 但是因为是 / 他们要海运 // 所以比较长的时间 // 大概一个月以后 // 才可能运到 [681698](#) 这个现在呢 // 中国也出了一些针对西班牙语国家的教材 // 就是它的注释翻译都是用西班牙语的 [690301](#) &mm 其中我知道有一个比较好的教材 // 我觉得它很口语化的呢 // 就是一个北大出的

教材 // 叫博雅汉语 ///α<701307> 那个呢 // 它是英文注释 ///α<703233> 它是从最基础的零起点 // 一直到最高级的 &mm 这个水平都有 ///α<711538> 这 [/] 这个教材呢 / 它对语法的解释比较简单明确 ///α<717443> &mm 而且它的这个语言呢 // 是很地道的简单的口语 ///α<723175> 这个我觉着对 [///] 特别是对初学者来说 // 那本书是不错的一本书 ///α<728412> &mm 还有呢 // 如果等我的这个教材送来以后 // 还有包括一些光盘 // 我可以给大家看一下 ///α<740478> 然后大家 / &mm 如果有兴趣的话 // 可以想办法买到它 ///α<745817> 对 // 因为很多同学要去中国 // 以后好像我们系有这种计划 // 就是说你四年的学习当中 // 至少要有的一年去中国留学的时间 ///α<756476> 这种 [/] 这种情况呢 // 我就需要 [/] 我就希望大家能够在中国的时候 // &mm 尽量地不 [/] 不过多地跟说西语的人来接触 // 而是尽量地跟外国人 // 就是西班牙语 [///] &mm 不是 [///] 母语不是西班牙语的这个 / 人来接触 ///α<779469> 因为你在中国的那个环境中 // 你上课的班里的同学可能很多都 [/] 都是来自不同国家的 // 那么多跟他们接触 ///α<786829> 还有呢 / 就是多跟中国人接触 ///α<789187> 你去商场 // 你去买 [/] 超市买东西 // 你尽量地用你学过的语言 // 学过的汉语来表达 ///α<797341> 这是一方面 ///α<798641> 还有呢 // 就是你要 [/] 就是尽量地跳出你西班牙语的这个生活圈子 // 来多接触其他的语 [///] &mm 其他国家的 // 用汉语来 [/] 来进行交 [/] 交际 ///α<812046> &mm 至于这个生活习惯方面呢 // 确实中国跟西班牙有很多的不同 ///α<819686> 比如说 / &mm 当然最 [/] 最常见的 // 用筷子 / 用叉子的这个不同 ///α<825415> 还有呢 / 吃饭的这个风 [/] 习俗 ///α<827833> 比如说 &mm 这个中国人都是一起来吃饭 // 我们一个大的盘子里的菜 // 我们 &mm 大家一起吃 ///α<837438> 还有呢 // 有一个很明显的不同 ///α<840096> 在西班牙的时候 // 可能吃饭的时候 // 人们吮手指 / 舔手指 / 这是很正常的 // 但是 // 在中国 / 如果你这样做呢 // 可能就会 [/] 别人就会觉得不习惯 ///α<851682> &mm 就是说这些 // &mm 餐桌上的这个 // &mm hhh {%act: clearing of throat} 一些习惯 ///α<861345> 再有呢 // 就是 / 比如说在饭馆儿里 // &mm 在饭馆儿里有什么不同 ? α<868803> 点菜 ///α<869551> 对对 hhh {%act: assent} // 这个是需要注意的 ///α<871748> 一个是水 // 我们的自来水是从水龙头里出来的水 // 是不能直接喝的 // 需要烧开 ///α<879103> 或者是呢 // 你要买那种 [///] 从超市里买成瓶的矿泉水 ///α<883856> 还有一个呢 / 就是很多洗手间 // 可能没有卫生纸 // 也需要你自己带 ///α<891143> 这个要注意 ///α<893405> 哦 / 对 // 中国的这个公共汽车什么都是有售票员的 ///α<901817> 一般来说是需要你买票/// 而且他是根

据你坐的这个距离的长短 // &mm 来收不同的钱 ///α<911909> 价格不一样 ///α<914164> 就是说强调 [/] 我们强调听说读写 // 这四个方面 // 但是实际上 // 最重要的还是听和说 ///α<923701> 所以上课呢 / 就 // 最强调的就是开口率 // &mm 强调精讲多练 ///α<931068> 老师少讲 // 然后学生多说 ///α<933861> 但是对于初级班来说 // 汉字也 [/] 也非常重要 ///α<937938> 汉字的认读和 [/] 和书写也非常重要 ///α<942780> 好 // 我们来看 [/] 我们来看第 15 课吧 hhh {%com: open the book} // 今天我们学习第 15 课 ///α<953850> 那么这一课书中 / 最重要的一个语法就是情态补语 ///α<960156> 这个补语对外国学生来说 // 是比较难的 / 一个语法 ///α<966827> 因为很多语言当中 // 没有这个补语 ///α<970421> 比如说 / &mm 我八点来学校了 ///α<975138> 来得早吗 ? α<976847> 来得很早 ///α<979115> 这个得 // 就是来得早中这个就是我们所说的 // 补 [/] 情态补语 ///α<985904> &mm 那么第二个 [///] &mm 还有一个方面呢 // 就是说 / 在用 =α<996467> 这是第一个补语的用法 // 情态补语第一个用法 / 就是没有宾语 // 就直接说 // 来得很早 ///α<1003761> 今天你几点来学校的 ? α<1006093> 今天我七点就来了 ///α<1008229> 他 [///] 你来得很早 ///α<1010333> 那么第二种情况 // 还有一种 [/] 一种情况 // 就是 // &mm 上海人说普通话说得好不好 ? α<1019256> 那么我们说 // 上海人说普通话 // 那我用情态补 [///] 用这个补语的时候 // 我们必须再说 // 上海人说普通话说得很好 // 必须再说一遍这个说 ///α<1032181> 要重复一遍这个 [/] 这个动词 ///α<1034581> 另外呢 / 我们也可以这样说 // &mm 普通话上海人说得很好 // 把这个普通话这个词呢 // 放到前边去 ///α<1044126> 这个时候 // 我们就不需要再重复这个动词 ///α<1047459> 这是第一个重点语法 ///α<1049815> 第二个语法呢 / 是 / 了 ///α<1052031> 这个了呢 // 是表 [///] 是放在句子的最后边 ///α<1055605> 表示以前 / 做了一个事情 // 这个事情已经完了 // 这个事情已经完成了 // 做完了 ///α<1063693> 注意 [/] 需要注意的就是说 // 这个了呢/// 它一定要放在句子的最后 ///α<1070081> 以前我们学过的这个了呢 // 是放在动词的后面 ///α<1074198> 比如说 / 我看了 [/] 昨天我看了一个电影 ///α<1080969> 这个了呢 / 是放在动词的后面 // 昨天我看了一个电影 ///α<1085251> 但是我们今天要学的这个 [/] 这个了呢 // 这个是放在句子的最后 ///α<1092463> 比如说 / 他去上海了 ///α<1095055> 他去上海了 // 那么这个句子的意思呢 / 就是他已经去上海了 // 这个事情已经完了 // 他去这个动作已经做完了 ///α<1103881> 第三个语法 // 这个比较简单 // 就是 // 这个 &mm 动词的重叠 ///α<1116288> 比如说我们说 // 学习 [///] &mm 休息 // 休息 // 我们可以说 // 我累了 // 我想休息休息

///α<1123999> 这个动词的重复呢就表达 // 时间很短 // 或者是不是很大的 // 不是很重要的一个事情 // 很简单的一个事情 ///α<1134785> 这个时候 // 我们可以用重复这个动词 ///α<1137535> 再比如说 // &mm 这个菜很好吃 // 你尝尝 ///α<1142288> 那么这个尝尝呢 / 就是 / 表示试一试的意思 ///α<1147605> &mm 我再说说关于发音的问题 ///α<1157338> 汉语的发音呢 // 对 / &mm 西班牙的同学来说 // 是很难的 ///α<1162546> 我觉得最难的就是它的声调 ///α<1164516> &mm 这个 āáàà 四个声调当中 // 尤其难的是二声和四声 ///α<1171526> 也就是二声上不去 ///α<1173481> 总是 / 比如说 // 你好 ///α<1178378> 这个第一个你变成了你好 // 常常说你好 // 这个声音上 [//] 二声上不去 ///α<1186281> 还有呢 // 就是四声下不来 ///α<1188848> 四声呢 / 读得也比较平 ///α<1194186> 这个是一个 &mm 主要问题 // 就是二声要往上去 // 而四声呢 / 要下降 ///α<1202544> 比如说 // 我从学院 // 去邮局 ///α<1209662> 这个邮局两个字都是 // 二声 ///α<1214555> &mm 他从英国 / 来中国 ///α<1219459> 来中国 / 特别是二声和四声放在一起两个词 // 在一起的时候 // 可能就会出现这样的问题 // 就是读得不准确 ///α<1234950> 比如说 / 学校 // 一个二声 / 一个四声 ///α<1244464> 那么这个需要大家 / 就遇到这种词的时候 // 要特别的注意 ///α<1250470> 学校 // 二声要提上去 // 而这个四声呢 / 要降下来 ///α<1255978> 这是一个声调方面的问题 ///α<1258914> 再有一些 [//] 一个问题呢 // 就是有一些发音 // 在西班牙语里是没有的 ///α<1264436> 比 [j] 比如 // 就是一些爆破音 ///α<1268364> 比如说 B 和 P // G 和 K 这样的 ///α<1274694> G 和 K // 这个 K // 要有气 [j] 气流来 [j] 出来 ///α<1280734> 这个时候你要用一点儿力气 ///α<1282789> 这个两个 [//] 这 [j] 这些音的这个区别 ///α<1288367> 还有就是 ZCS 和这个 JQX 这几个音 ///α<1296741> JQX // 这几个音 // 是不是在西班牙语也没有 ? α<1307641> JQX 也没有 ? α<1309368> 还有 ZCS / 也没有 ? α<1311662> &mm 所以这几个音呢 / 学习起来是比较困难的 // 需要大家多听 ///α<1318393> 听了以后呢 / 模仿 // 要 XXX 纠正你以前说西班牙语的一个 [//] 一些发音习惯 ///α<1327800>

## 2. Ejemplo de subcorpus Informal

[  
@Title: mascota  
@File: chin13

@Participants: JIJ, (woman, A, 3, student, participant)

ZHF, (man, A, 3, student, participant)

@Date: 01/05/2009

@Place: Alcala de Henares

@Situation: in a room, not hidden

@Topic: Rabbits, dogs, cats, ducks

@Source: C\_ORAL\_CHINA

@Class: informal, public, dialogue

@Length: 16' 20"

@Characters: 4753

@Acoustic\_quality: A

@Transcriber: Dong Yang

@Revisor: AZY

@Comments:

]

\*JIJ: 哎 / 你养了一只兔子 // 是吧 ? α<2067>

\*ZHF: 是啊 hhh {%act: assent} // 你怎么知道的 ? α<3639>

\*JIJ: &mm 地球人都知道 ///α<5733>

\*ZHF: hhh {%act: laugh} ///α<5992>

\*JIJ: 那 [/] 那个兔子是什么种类的 ? α<7961>

\*ZHF: &mm 安哥拉兔 ///α<9101>

\*JIJ: 什么样子呢 ? α<10425>

\*ZHF: &mm 白白的 // &mm 耳朵也比较长 // 但是比其他的那种兔类 // 看起来 // 就要短些啦 // 因为它的毛比较长 ///α<17118> &mm 看起来像一个 // &mm 大毛球一样 // 是白色的 ///α<21293> 然后 // 还是笨笨的样子吧 ///α<23505>

\*JIJ: 那肯定很可爱 // 手感肯定不错 ///α<25689> 那养兔子麻烦吗 ? α<27147>

\*ZHF: &mm 其实不如我想象中的麻烦 ///α<30062> &mm 因为一开始 / 买的时候 // 他们会说养兔子会有很多问题 ///α<34202> 但是最后 / 看起来的话 // 其实每天给它收拾两次左右吧 // 就能基本保证 / 这个它的房间的卫生了 ///α<41471>

\*JIJ: 那它吃什么呢 ? α<43055> 很麻烦吗 ? α<43882>

\*ZHF: &mm 我的兔子是什么都可以吃 ///α<46621> 它 [///] 一开始 // 因为兔子比较小吗 ///α<49113>



\*JIJ: <hhh {%act: assent}>

\*ZHF: [<] <所以说> / 喂养的时候 // 比较用心 ///α<51261> 然后要注意给它的饮水 // 不能给它喝生水 // 要让它喝矿泉水啊 / 纯净水啊 ///α<57001> 然后 // 吃的方面 // 只能给它喂那个 / 精装的兔粮 ///α<61726> 但是后来等兔子到三个月以后 // 然后 / 我就开始试着给它喂各种各样的东西 ///α<66792> 比如说 / 水啊 // &mm 比如说胡萝卜啊 / 苹果啊 // 然后各种蔬菜 ///α<72800> 然后它特别喜欢那种 / 咬起来口感有点儿硬 // 而且口感比较甜的食物 // 比如说 / 胡萝卜 // 这是它最喜欢的/// 然后苹果也很喜欢 ///α<81075> 其他的 / 像蔬菜 // 豆类 // 然后 // &mm 玉米 // 马铃薯 // 土豆儿 / 都可以吃 [/] 噢不 / 马铃薯之类之类都可以吃 ///α<93125>

\*JIJ: 你实在太专业 / 太体贴了 // 你们家兔子肯定很幸福 ///α<96494> &mm 那 // 我记得上次 / 我给它喂了块苹果 // 它都不吃了 ///α<101285> 就是被你喂馋了吧 ? α<102640>

\*ZHF: <&mm>

\*JIJ: [<] <你太宠着它了> hhh {%act: laugh} ///α<104437>

\*ZHF: &mm 这倒不是的 ///α<105415> 因为上次你过来的时候 // 是 / 下午的时候 ///α<109179> 然后我们众所周知 // 兔子跟我们人类的时间差 // 是很大的 ///α<112447> 一般它的活动时间是在晚上十点之后 ///α<114947>

\*JIJ: &mm ///α<115089>

\*ZHF: 有次 [/] 上次你过来的时候 // 它正在 [/] 相当 [/] 相当于它正在睡觉 ///α<118659> 然后 // 先 [///] &mm 比较难跟它交流 ///α<121831> 但是一般到晚上三四点钟的时候 // 它就会特别活跃 // 这个时候是它正常的进餐时间 ///α<126866>

\*JIJ: 兔子叫什么呢 ? α<128656>

\*ZHF: &mm 是西语名还是中文名 ? α<131089>

\*JIJ: 当然是西语名了 ///α<132289>

\*ZHF: &mm 它叫 LECHUGA ///α<133517>

\*JIJ: &mm 那它现在认识你吗 ? α<136283>

\*ZHF: &mm 我觉得它现在只认识胡萝卜 // 并不认识它的主人 hhh {%act: laugh} ///α<140726>

\*JIJ: 我就觉得像兔子这种脑袋比较小的动物 // 都是脑容量不足 // 所以它肯定不认识主人 ///α<146900> 这就跟狗不一样 ///α<148484> 像狗 / 头那么大 // 所以它就好比兔子要聪明很多 hhh {%act: laugh} ///α<151615>

\*ZHF: 是吗 / 你原来难道养过狗 ? α<153526>

\*JIJ: 我没有养过狗 // 但我特别喜欢狗 ///α<155582>

\*ZHF: 是吗 ? α<157168>

\*JIJ: 因为我妈不准我养狗 ///α<158931> 她就是 // 所有带毛的动物她都害怕 ///α<161510> 所以我就不能养狗 // 但是我特别喜欢狗 ///α<163825>

\*ZHF: hhh {%act: laugh} 你为什么不养一只乌龟呢 ? α<165189>

\*JIJ: 我不喜欢乌龟啊 ///α<167388> 因为我 [/] 我不喜欢那种就是不认识主人的 // 然后头特别小的动物 ///α<171518> 像狗就脑容量比较大 // 就比较聪明 // 所以我比较喜欢 ///α<175278> 而且像那种大狗 / 让人觉得比较有安全感 // 又毛毛的 // 摸上去比较舒服 ///α<180334>

\*ZHF: &mm 你刚才提到了大狗 ///α<181703>

\*JIJ: <&mm>

\*ZHF: [<] <那> 你最喜欢哪种大型犬啊 ? α<183583>

\*JIJ: 哈士奇 / 哈士奇 hhh {%act: laugh} ///α<184999>

\*ZHF: 哈士奇是吗 ? α<185823>

\*JIJ: 雪橇犬 ///α<186570>

\*ZHF: 就是那种灰白相间的 ///α<188187>

\*JIJ: 对 hhh {%act: assent} 有蓝眼睛的那种 ///α<190226>

\*ZHF: 噢 / 它是产自哪个国家的呀 ? α<192250>

\*JIJ: 爱斯基摩人的吧 hhh {%act: laugh} // 我不知道 ///α<195830> 因为它本 [//] 它不是又叫哈士奇又叫雪橇犬 ///α<199291> 所以本来是用来拉雪橇的 ///α<201277> 可能就是在很寒冷的地方 // &mm 用来拉雪橇的那种狗 ///α<206022> 很可爱 / 很可爱 ///α<207427>

\*ZHF: &mm 确实狗是我们人类最好的伙伴 hhh {%act: laugh} ///α<210361>

\*JIJ: 因为第一次见哈士奇 // 是在谢霆锋一张专辑的封面上 ///α<214353> 他放着那只哈士奇 hhh {%act: laugh} // 我就觉得特别可爱 ///α<217376> 然后后来 [/] 后来又看了一个电影 // 叫南极大冒险 ///α<220246> 就是讲 // 好多只哈士奇一起在 [/] 就是南极的那种历险故事 // 就很好玩儿 ///α<226651> 然后咱们学校以前有好多只哈士奇 // 我都抱过 // 很好玩儿 ///α<231362>

\*ZHF: &mm 你说到咱们学校 // 我想起来了 // 原来咱们学校还有不少野猫 ///α<234357> 你还记得它们吗 ? α<235542>

\*JIJ: 当然了 // 有时候还去喂它们 ///α<238677>

\*ZHF: 是吗 ? α<239138> 我记得有一只叫大黄的 // 然后胖胖的 // 已经被我们学校的女孩子宠得很挑剔了 ///α<244254>

\*JIJ: 噢 / 那只大黄就很像加菲 ///[246849](#)> 长得很像加菲猫是  
吧 ? [249246](#)>

\*ZHF: <&mm>

\*JIJ: [<] <就> 肥肥的那个 ///[250192](#)>

\*ZHF: 对 hhh {%act: assent} ///[250748](#)>

\*JIJ: 简直就是明星脸 ///[252264](#)>

\*JIJ: 对啊对啊 ///[253075](#)> 他们说 // 那只猫 // 现在 / hhh {%act:  
click} 由于我们学校的 / 那个同学们比较友爱 // 已经连 / 火腿肠都不  
吃了 ///[260308](#)> 要给它吃鱼香肉丝之类的成菜 // 它才有胃口  
///[263464](#)>

\*JIJ: 看来动物的命运也真是 / 各不相同 ///[266151](#)> 有的动物这  
么幸福 // 有的动物那么可怜 ///[269585](#)> 有的猫连水煮鱼都吃过  
///[272340](#)>

\*ZHF: hhh {%act: laugh} 就是就是 hhh {%act: assent} ///[273450](#)>  
那你原来家里养过什么动物吗 ? [275165](#)>

\*JIJ: 我养过小鸡 / 小鸭 / 小鱼 // 还有 [//] 反正都是 / 就没有一  
个能活得长的 hhh {%act: laugh} ///[283332](#)>

\*ZHF: 鸭子看起来是很可爱的动物呢 ? [285647](#)>

\*JIJ: 没错 // 但是是被我淹死的 hhh {%act: laugh} ///[288073](#)>

\*ZHF: hhh {%act: laugh} // 淹死的 ? [289494](#)>

\*JIJ: 因为好像它是只旱鸭子 // 那我把它抱到水里面 // 让它游泳 //  
然后它就 / 可能最后游不动了 // 就 [/] 就不行了 ///[296154](#)>

\*ZHF: 是啊是啊 // 我家原来也养过鸭子 ///[297850](#)> 然后 / 是一  
只很贪吃的鸭子 ///[300235](#)> 然后 / 周末的时候 // 我会和我爸爸妈妈  
一起去公园里 // 给它捞鱼 ///[304857](#)> 用网兜 / 网小鱼给它吃 // 它特  
别喜欢 ///[307724](#)>

\*JIJ: 我觉得你 // 你对宠物都特别好 // 特别有爱心 ///[312369](#)>  
很专业 ///[314231](#)>

\*ZHF: 谢谢谢谢 ///[315370](#)> &mm 我们家原来还 [/] hhh {%act:  
click} [/] &mm 我们 [/] 我们家原来养过兔子 ///[319201](#)> 就是 [//] 但  
是是一只野兔 ///[321381](#)>

\*JIJ: <&mm>

\*ZHF: [<] <那时候> 是我跟我爸爸去打乒乓球 ///[323897](#)> 在一个  
乒 乓 球 =[325043](#)> 我们因为当时是是在一个学校大院里打  
///[327689](#)> 然后在乒乓球台下面发现它的 ///[329859](#)> 然后 // 是一只  
比较野性难驯的兔子吧 ///[333318](#)> 来我们家一开始 // 两个 [/] 两个多

月基本上不太吃东西 // 然后奄奄一息的样子 ///α<338634> 后来时间长了 // 相处得比较熟了 // 然后 // &mm 它才开始渐渐地吃东西 // 开始渐渐地阳台上活动起来 ///α<346957> 最后还做了一个小草窝 ///α<348804>

\*JIJ: &mm 很好 ///α<350198> 你知道吗 ? α<350813> 那个 / 其实最烈性的动物是 [/] 是那个麻雀 ///α<354986> 就是它永远不会被人驯服 ///α<357005> 如果你把一个麻雀弄 [/] 弄过来养的话 // 它就会绝食 // 直到 [/] 直到死 ///α<361221> 因为我小时候有一次下大雨 // 住平房 // 然后我爸就在那个屋檐底下 // 就捡了一只小麻雀回来 ///α<368518> 但就养到第二天 // 想等它好点把它放走 // 结果它就死了 ///α<372142> 好像是自己咬断舌头自尽了 ///α<374199> 因为麻雀就特别有性格 ///α<376354>

\*ZHF: hhh {%act: surprise} 噢 // 还有这 [/] hhh {%act: laugh} 这么有性格的动物 ///α<380467>

\*JIJ: 就是我 [/] 我也很奇怪 // 就后来看一个 [/] 看篇文章才知道 // 原来麻雀这么烈性 ///α<386503>

\*ZHF: &mm ///α<389100>

\*JIJ: 但是我最喜欢的动物还是狗 ///α<392744>

\*ZHF: 是吧 ? α<393621>

\*JIJ: &mm hhh {%act: assent} ///α<394125> 你喜欢小狗吗 ? α<395355>

\*ZHF: &mm 挺喜欢的 ///α<398349> 然后 [///] 但是 // 因为狗的种类有很多吗 ///α<401822> &mm 要 [///] 我比较喜 [/] 我比较喜欢 =α<406846> 我都不知道 // 有一种叫吉娃娃的狗 // 你知道吗 ? α<409583>

\*JIJ: 知道 ///α<410529> 我不喜欢那个狗 ///α<411552>

\*ZHF: 对啊 // 我觉得很可爱的 // 长的像 ET // 长得像外星人一样 ///α<415038>

\*JIJ: 但是那种小狗 // 就是你让它放在家里面跑 / 一不小心 // 就会踩着它 ///α<419672>

\*ZHF: hhh {%act: laugh} ///α<419821>

\*JIJ: 太可怕了 ///α<420798>

\*ZHF: 噢 / 对 // 你听说过吗 ? α<422309> 一般狗越小脾气越大 ///α<424428>

\*JIJ: 没错 ///α<425432> 像那种大狗 // 一般你去摸它 // 它都没反应 // 都很好 // 很温顺 ///α<428937> 然后有时候小狗 // 你一上前 // 它

就会乱叫 ///α<432177>

\*ZHF: 对对对 hhh {%act: assent} ///α<433573>

\*JIJ: 我有一次在那个 / 银行门口 // 看一只狗被栓在那儿  
///α<437445> 它主人可能进去取钱了 // 它在那栓着 ///α<439800> 然后它  
嘴上就戴着那个 [ ] 就是那个 / 护套还是什么 // 就是防止它咬人的那种  
东西 ///α<445473>

\*ZHF: &mm ///α<445756>

\*JIJ: 我当时就觉着它很可怜/// 因为那么小 // 又没什么攻击性 //  
又戴上那个 ///α<449709> 然后后来我就想 / 想过去摸摸它 ///α<452347>  
然后我 / 还 [ ] 就刚走了一步 // 它就冲我特别凶地叫 ///α<455594> 然  
后我终于知道为什么它要戴那个护套了 hhh {%act: laugh} ///α<458104>

\*ZHF: &mm 但是我觉得 // 在养动物的过程中 // 我们还是应该比  
较尊重动物的天性的 ///α<462494>

\*JIJ: &mm ///α<462711>

\*ZHF: 比如说 / 原来有的人养狗的时候 // 他会给狗 / 戴上嘴套或  
者头套 ///α<466930> 就是因为狗会叫 // 然后叫的话显得比较不乖吗 //  
有失主人的身份 ///α<470990> 但是长此以往的话 // 这个狗就会特别胆  
小 // 非常怯懦 ///α<474172> 然后跟同伴 // 就是它在街上遇到别的狗的  
时候 // hhh {%act: click} 之间交流也会有问题 ///α<478740> 然后 // 这时  
候我觉得狗 // 应该 // 相当不开心的 hhh {%act: laugh} ///α<483750>

\*JIJ: &mm 没错 ///α<484587> 还有 // 我觉得狗本来就是应该是家  
养动物 // 因为是从狼驯化来的 ///α<489179> 所以就是 // 有时候看到街  
上那些流浪狗 // 就觉得它们很可怜 ///α<493337> 因为它们已经失去了  
动物的天性 // 没办法自己在野外生存 ///α<496956> 就那些人太不负责  
任了 ///α<498515> 你既然把它驯养了 // 就不应该抛弃它 ///α<501489>

\*ZHF: &mm ///α<502075>

\*JIJ: 因为那只是你生命中的几年 // 但是 / 是小狗的一辈子  
///α<505315>

\*ZHF: 对对对对对 hhh {%act: assent} ///α<506464> &mm 你这么喜  
欢狗 // 你喜欢猫吗 ? α<509638>

\*JIJ: 不喜欢 ///α<510213>

\*ZHF: 为什么呀 ? α<510970>

\*JIJ: 因为猫头很小 hhh {%act: laugh} // 所以脑容量不足 // 跟兔子  
是一个原理 hhh {%act: laugh} ///α<514585>

\*ZHF: 看起来你喜欢各种各样头大的动物 hhh {%act: laugh} // 是  
吗 ? α<517511>

\*JIJ: 不是头大 / 就是比较聪明的 ///α<519144>

\*ZHF: 认主人 ? α<519890>

\*JIJ: 对 ///α<520237> 我觉得猫有时候 // 就是 [///] 反正不是很 [/] 不是很认主人 // 而且比较冷 // 不像狗那么忠心 ///α<527443>

\*ZHF: &mm 有一种说法 // 说狗比较像男性 // 比较忠诚 // 然后猫比较像女性 // 比较有性格 ///α<533667> 然后 / 有时候你冷落它 // 它 [/] 它也会不理你 ///α<536690> 然后 // &mm 比较难以猜测它的心思吧 ///α<541620>

\*JIJ: &mm 我觉得 / 就猫 / 就是那种 / 没有表情的动物 // 但是狗就是有表情的 ///α<548091>

\*ZHF: 狗是有表情的 hhh {%act: surprise} ? α<550093>

\*JIJ: 我觉得是 // 因为它 [/] 它要是难过 // 它会 // 就那种呜呜地叫 ///α<554945> 然后 / 开心的话 // 就会 [///] 我觉得狗是会笑的 ///α<558500>

\*ZHF: <&mm>

\*JIJ: [<] <我不> 知道你能不能看出来 hhh {%act: laugh} ///α<560797> 反正这是 / 我的感觉 hhh {%act: laugh} ///α<562676>

\*ZHF: 是吗 ? α<563470> &mm 那根据你的了解 // 狗一般是怎么跟人类交流的呢 ? α<567751>

\*JIJ: 就是 // 用 // 摇尾巴呀 // 来表示友好 ///α<572543> 然后 / 用一些就各种各样的叫声 // 来表达它自己的想法 ///α<576993> 然后 [///] 或者可以拖着你玩啊 // 什么的 ///α<579787> 因为我没养过狗 // 所以我不了解 // 我都是到处找别人的狗玩儿 ///α<583834>

\*ZHF: hhh {%act: laugh} 其实 / 很多动物都有自己的语言 ///α<588229> 比如说 // 蜜蜂发现食物的时候 // 它会圆圈 [///] 原地跳八字舞 ///α<592861> 然后当兔子感觉危险的时候 // 它会不断蹬后蹄发出响声 // 来警告自己的同类 ///α<597699> 然后虽然有些宠物 / 现在已经家养了 // 但是它们原始的天性还是得以保留的 ///α<602508>

\*JIJ: &mm 很有意思 ///α<604354>

\*ZHF: &mm 你知道藏獒吗 ? α<610719>

\*JIJ: &mm 知道 // 很有个性的狗 ///α<612990>

\*ZHF: &mm 在我国的西北部 // 藏獒有 [/] 有一种说法 ///α<617500> 叫做 // &mm 一獒可敌三狼 ///α<620493> 就是说獒是一种很忠心 // 同时很勇猛的动物 ///α<623814> 它们只认识从 [/] 从小生下来就照顾它们的那些主人 ///α<627679> 然后 / 很难跟它们建立关系 // 但是 / 你一旦跟它认识之后 // 它就对你 [/] 会对你特别忠诚 ///α<632699>

然后 / 是享誉我 [/] &mm 海外 [/] 享 [/] 享誉国内外的一种很有名的犬类 ///α<638522>

\*JIJ: 而且很臭 ///α<639446> 因为 / hhh {%act: laugh} 我有一个好朋友他舅舅就专门养大型犬的 ///α<643534>

\*ZHF: &mm ///α<643936>

\*JIJ: 然后他就 [///] 他家后院儿一个 [///] 就好多铁笼子 // 然后有几十条那种大狗 // 就藏獒还有其他别的狗 // 我忘了 ///α<650854> 反正就 / 真的超级超级臭 // 臭得难以想象的那种 ///α<655644> 然后而且很凶狠凶 ///α<657686> 但是它们对它的主人就特别忠心 // 特别好 ///α<660889>

ZHF: &mm ///α<661564>

\*JIJ: 而且我觉得现在藏獒因为 / 越来越多人把它作为家养的宠物 // 它有 [/] 已经有点儿失去它原来的天性了 ///α<667855> 现在一只家养的藏獒肯定敌不过三头狼 ///α<670482>

\*ZHF: hhh {%act: laugh} 是啊是啊 hhh {%act: assent} ///α<672248>

\*JIJ: 其实我觉得 =α<674745> 你 [/] 你喜欢狼吗 ? α<676598>

\*ZHF: &mm 我不喜欢 // 但是我比较欣赏这种动物 // 因为原来看过一本书 / 叫狼图腾 ///α<681740>

\*JIJ: &mm ///α<681872>

\*ZHF: 我已经记不清作者了 ///α<682948> 但是当时是正 [/] 我正在上高中的时候 ///α<685693> 但这本书 / 给我留下了很深刻的印象 ///α<688060> 它向我们诠释了一些概念 ///α<690103> 然后 // hhh {%act: click} &mm 当时是 [/] &mm 当时对我影响很大 ///α<694545> 主要是因为当时正在高考吗 ///α<696558> 然后 // 人与人之间的竞争比较激烈 ///α<699338> 然后那本书就向我们诠释一种狼的精神 ///α<702017> 比如说在受伤的情况下不断拼搏 ///α<704897> 在失去同伴的情况下 // &mm 依然能够 [/] &mm 能够自己独立地生存 ///α<711143> 然后 [///] 但是同时它又非常依赖于群体 // 然后对群体有很高的信任 ///α<716295> 有些狼甚至会牺牲自己的性命 // 来保全这个团 [/] 团队的利益 ///α<720126>

\*JIJ: &mm 我记得有个说法是 // 如果一只狼 / 被兽夹 / 夹住了脚的话 // 它为了 [/] 为了生存下去 // 它会自己咬断自己的骨头 ///α<728146>

\*ZHF: 对对对 hhh {%act: assent} ///α<729542>

\*JIJ: &mm 我 [/] 我其实挺喜欢狼的 // 但是是那种叶公好龙式的喜欢 ///α<735147>

\*ZHF: &mm ///α<735370>

\*JIJ: 就是离的远远的 ///α<736634> 我还觉得狼 ///α<737359> <狼是挺不错的>

\*ZHF: [<] <hhh {%act: laugh}> ///α<738554>

\*JIJ: 因为那个 [/] 因为它的外形就 / 有点像哈士奇 ///α<742085>

\*ZHF: &mm ///α<742748>

\*JIJ: 就是那 [/] 那种样子我挺喜欢的 ///α<744617> 然后以前还梦想过要养只狼或 <养只豹子什么的>

\*ZHF: [<] <hhh {%act: laugh}> ///α<750207>

\*JIJ: 还可以保护我 hhh {%act: laugh} ///α<751474>

\*ZHF: 好像施瓦辛格养了一只老虎还是狮子 // 你知道吗 ? α<754661>

\*JIJ: 不知道 hhh {%com: move the head} ///α<756083>

\*ZHF: &mm 我也是听说吧 ///α<759169> 反正我觉得跟如此大 [/] 一个大型动物在一起 // 感觉一定会比较有意思 ///α<764490>

\*JIJ: 可是这个是合法的吗 ? α<766171> 在美国 ? α<767353>

\*ZHF: 不知道 ///α<768868> 我觉着 / 估计是类似 [/] 是类似一种领养的性质吧 // 就好像咱们国家大熊猫 ///α<773388> 由于经济问题 // 咱们 / 当地的 // &mm 有关机构不能够负担它的费用的话 // 他就会让一些人来领养大熊猫 ///α<780697> 然后通过这种方式 // 来让大熊猫获得更好的生存条件 ///α<784188>

\*JIJ: &mm 我突然想起来 // 就是前一阵子去希腊旅游 ///α<788194>

\*ZHF: &mm ///α<788431>

\*JIJ: 就发现希腊 // 就是尤其是雅典 // 到处都是流浪狗 ///α<792767> 但是那些狗都带着狗牌儿 ///α<794823> 然后是一样的狗牌儿 ///α<796120> 所以我就很奇怪/// 就还 [///] 一开始还以为是个政府雇狗来看城的 / 还是什么意思 ///α<801446> 然后后来那个旅店的人 / 就说那个就是流浪狗 ///α<804805> 然后我就不明白 // 因为 [/] 就是因为那些狗都是 [/] 都是那种大型犬 // 然后都带的是一样的狗牌儿 ///α<811581> 我就觉得那些主人实在是太狠心了 ///α<813925>

\*ZHF: 对啊 // 但是这些狗牌是谁给它们装上的呢 ? α<817552>

\*JIJ: 所以就 // 是个谜啊 hhh {%act: laugh} ! α<819437>

\*ZHF: hhh {%act: laugh} 都是一样的狗牌儿 // 但是狗牌儿上有不同的名字还是 .....α<824432>

\*JIJ: 没细看 // 不敢接近 hhh {%act: laugh} ///α<826346>

\*ZHF: &nenenenenenene ///α<827771>

\*JIJ: 但到处都是那样的狗 ///α<829631> 就那些历史遗迹里面也都



是 ///α<832970>

\*ZHF: 是吗 ? α<833798>

\*JIJ: &mm 然后就像给 [//] 就是 [//] 就像守城的狗一样  
///α<842205> 还有咱们在意大利那个庞贝古城里面 // 不也有狗吗 ?  
α<846312>

\*ZHF: &mm 我觉着那些好像是那个 / 当地的政府专门 / 雇佣过来  
// 然后帮 [/] 帮助保护文物古迹的 ///α<853528>

\*JIJ: 我有时候在想 // 会不会是以前那些狗的灵魂 ? α<857205>

\*ZHF: hhh {%act: laugh} ///α<858476>

\*JIJ: 就是 // 那些人不在 / 但那些狗还守护在那个城市 // 所以  
在那徘徊来徘徊去的 ///α<864852>

\*ZHF: 你说的这个让我想起了赤兔马的故事 ///α<867310>

\*JIJ: &mm ///α<868786>

\*ZHF: 当时不是那个叶公 {%alt: el dato correcto es 关公} 走麦城之后  
// 然后赤兔马 // 三日不吃不喝 ///α<874159> 然后 // 为离开自己的主人  
吗 ///α<877334> 然后 // 最终 / hhh {%act: click} 也算是成全了它主人  
的一段英明吧 ///α<881457>

\*JIJ: 没错 // 马就是除了狗以外 / 我另外喜欢的一个动物  
///α<885185> 就是很有灵性 // 很可靠 // 而且很实用 ///α<888559>

\*ZHF: 你骑过马吗 ? α<889575>

\*JIJ: 骑过 ///α<890264>

\*ZHF: 在内蒙古 ? α<891484>

\*JIJ: 没错 ///α<892125>

\*ZHF: 感觉怎么样 ? α<893515>

\*JIJ: 很爽 hhh {%act: laugh} ///α<894573>

\*ZHF: 那你是怎么跟马交流的呢 ? α<896855>

\*JIJ: 那个时候 // 很害怕 // 就坐着一动都不敢动 // 没怎么跟它交  
流 ///α<901876> 下次有机会再好好交流 hhh {%act: laugh} ///α<904193>

\*ZHF: hhh {%act: laugh} 你骑的那种马是有马鞍的吗 ? α<905985>

\*JIJ: 应该是吧 // 那种给游客骑的马都是备马鞍的 // 然后还有人牵  
着 ///α<911045>

\*ZHF: 我觉着纵马内蒙古一定是一件特别有意思的事情  
///α<915101>

\*JIJ: &mm 我觉得如果 // 有 [/] 有一个自己的农场 // 然后养好多  
狗 // 再养好多 [/] 好多匹马 // 然后人生就完美了 ///α<923513>

\*ZHF: hhh {%act: laugh} 很不错的理想呢 ///α<925828>

\*LIJ: 那个 # [///] 那你养过小狗吗 ? α<933521>

\*ZHF: &mm 没有 // 我一直挺遗憾的 ///α<935383> 因为 / 有很多朋友都在养狗/// 然后他们讲狗的各种各样有趣的故事 ///α<940378> 讲狗的忠诚啊 // 讲 / 小狗有很可爱的一面啊 // 之类之类的 ///α<945963> &mm 但是一直没有机会养 ///α<948666> 因为在家里毕竟条件比较有限吗 ///α<950897> &mm 一方面是经济上的 // 买一只狗在我们国内 [///] 买一只比较好的狗在国内还是很贵的 ///α<957546> 另外我们需要不断地为它准备饮食 // 准备各种住宅 ///α<962123> 然后帮它挑选 // hhh {%act: click} 同伴或者配偶 ///α<965113> 然后 [///] 另外一方面就是时间问题 // 因为咱们 // 国内大多数家长还都是上班族吗 ///α<970136> 然后每天都有工作要忙 // 晚上回来一般都比较疲劳了 ///α<973378> 很难保证小狗一日三餐的供应 ///α<975951>

### 3. Ejemplo del subcorpus de Medios de Comunicación

[  
 @Title: curriculum  
 @File: chme08  
 @Participants: LIJ, (woman, A, 3, student, participant)  
 @Date: 05/05/2009  
 @Place: Beijing  
 @Situation: in a studio  
 @Topic: curriculum vitae  
 @Source: C\_ORAL\_CHINA  
 @Class: formal, public, Media  
 @Length: 10' 03"  
 @Characters: 2469  
 @Acoustic\_quality: A  
 @Transcriber: Dong Yang  
 @Revisor: Yang He  
 @Comments:  
 ]

\*LIJ: {%com: music} ///α<11799> 观众朋友们 / 大家好 ! α<14293>  
 欢迎收看本期培训驿站 ///α<16523> 受金融危机影响 // 今年毕业生们在求职这条路上走得异常艰难 ///α<21901> 面对这样严峻的就业形式 // 以

及日益激烈的求职竞争 // 一份好的简历 / 对于成功求职 / 起着至关重要的作用 ///a<30908> 它甚至可以使你在上百名求职者中脱颖而出 ///a<35343> 据网络调查显示 // 28%的大学毕业生在写简历时 / 会求助于老师 ///a<41656> 9%的大学毕业生 / 会查阅相关书籍 ///a<44815> 而有 63%的毕业生 / 都是通过从网上下载模板来写简历 ///a<51010> 这类现象非常普遍 ///a<52872> 这就造成了 / 简历格式的公式化 ///a<56062> 往往招聘人员只会用 10 秒左右来考虑一份简历是否值得他继续读下去 ///a<63012> 而千篇一律的简历则不可避免地被淘汰 ///a<67074>

\*XYZ: 我是来自西葡语系的大四的同学 ///a<69617> 然后 / 现在刚刚签了一家国企 ///a<72494> &mm 一开始准备要写简历的时候 // &mm 感觉是无从下手 ///a<77895> &mm 也从网上搜了一些简历 // 借鉴了一些 [ ] 其他人的一些简历 ///a<82456> 发现这些简历嘛 // &mm 有出彩的地方 // 但是呢 / 放到我自己身上 // 有些地方还是 =a<88923> 怎么说呢 ? a<89566> &mm 没有办法突出我自己简历的优势 ///a<92549>

\*XYZ: &mm 现在大三了 // 正在打算 &mm 下学期 &mm 做一份这样的实习工作 ///a<99398> 我在大一大二还有包括大三上下学期 / 这三年当中呢 // 都参加了不少实践活动 ///a<106787> 然后 / 也积累了一定的工作经验 // 另外获得了不少的证书和奖状 ///a<112427> 但是我不知道在写简历的时候 // 我如何把我这些证书 / 奖状 // 把它们都归拢到一起 // 然后 / 来突出我的优势和强项 ///a<122812>

\*LIJ: 看来 // 要想做出一份好的简历 / 还真是不容易 ///a<126671> 下面 // 我们就给大家来介绍一下做简历的过程 ///a<130242> 首先 / 请大家来看一下简历的 3 种基本格式 ///a<134145> {%com: background music} 一 / 时序型简历 ///a<136119> 内容包括求职意向 // 教育背景 // 社会实践 // 校园活动和获奖情况等内容 ///a<145188> 时序型简历是指按倒叙的时间顺序 / 列出你的事业史// 在列出的每一项职位下 // 你要说明你的工作时限 // 你所承担的职责 // 该职位所需要的技能 / 和你曾经获得的突出成就等 ///a<161262> 时序型个人简历 / 比较适合于以下的几种情况 ///a<165642> 一 // 你申请的职位与你的教育背景和工作经历相符合 ///a<171264> 二 // 你有在著名公司工作的经历 ///a<174935> 三 // 你的工作经历能够反映出你的相关工作技能是不断提高的 ///a<180836> 四 // 你最近承担的职务 / 足以体现你的优势 ///a<185466> 二 // 功能型个人简历 ///a<188188> 包括求职意向 // 职业总结 // 职业经历等内容 ///a<193521> 功能型个人简历 / 主要强调你的技能 // 能力 // 资质以及成就 // 而不是你的事业史 ///a<201486> 它适合的情况有 // 一 // 跨专业求职 // 但你具有申请该职位所需的技能 ///a<208619> 二 // 你缺乏在

著名公司工作的经历 // 职位不高或者缺少荣誉奖励 ///α<215216> 三 // 你在短时间内换了很多不同种类 / 或者完全不相关的工作 // 但不想给招聘者 / 留下频繁更换工作的印象 ///α<225576> 三 // 混合型个人简历 // 包括职业总结 // 技能总结 // 职业经历 // 教育经历等内容 ///α<235318> 混合型个人简历是时序型个人简历与功能型个人简历的结合运用 // 可以按照时间顺序列出你的工作经历 // 教育背景等 // 也可以着重突出你的成绩 / 能力和优势 ///α<249465> 它适合的情况是/// 一 // 你想跳槽 ///α<252946> 二 // 你有过事业的巅峰 ///α<255319> 三 // 你既想要突出你的个人经历 // 又想要突出成就与能力 {%com: end of background music} ///α<260957> 以上就是我们在写简历时用到的三种格式 ///α<265372> 时序型简历突出你的经验和优势 // 功能型简历突出你的技能 // 混合型简历是以上两种的综合运用 ///α<273440> 在突出个人经历的同时 / 也着重描绘了成就与能力 ///α<277644> 当然 // 同学们也可以在此基础上进行发挥 // 做出适当的补充 ///α<282451> 现在 // 我们就来给大家介绍一下 // 一份出色的简历是怎样完成的 ///α<287592> {%com: background music} 如何制作简历 ///α<289332> 1 // 选择目标 ///α<291755> 如果你不明确自己想应聘的职位 // 可以先将拟定的几个目标罗列出来 // 逐一筛选斟酌之后 // 将它写在简历中 ///α<301751> 如果你已经明确了想要应聘的职位 // 也可以将它写在简历上 // 但一定要写清楚 ///α<309020> 2 // 列写教育程度 ///α<311480> 在已经选择的目标下面 // 列出你受过的相关教育和训练 ///α<316980> 持续的学习和训练说明你有上进心 // 所以 // 写出你在基本教育后所受的相关教育 // 这一点很重要 ///α<326587> 3 // 列写社会实践目录 ///α<329090> 把你参加过的社会实践都列出来 // 包括公司名称和地址 // 你所担任的职位和工作时间 ///α<337252> (1 // 适当地添加细节 // 详细描述 / 你在每段社会实践中所承担的职责 // 但是 // 应注意不要把它写成散文 // 或将职责的范围描述得过窄 ///α<350108> (2 // 准确地做出筛选 ///α<353173> 想想哪些实践经历与你的目标职位有关 // 删去无关紧要的部分 ///α<359561> 注意 // 简历是一张帮助你跨进企业门槛的名片 // 而不是工作回忆录 // 所以你只需要填写与所应聘职位相关的经历 ///α<370711> (3 // 着重叙述成功的实践经历 ///α<374548> 回顾刚刚列出的社会实践 // 想一想其中有没有你做的非常出色的案例 ///α<380831> 将叙述的重点 // 放在这些经历上/// 注意要多使用数字 // 因为数字有很强的说服力 ///α<388406> 4 // 修改与完善 ///α<391429> (1 // 添加关键字 ///α<393420> 现在许多人都用电子邮件或者可扫描的简历 ///α<398125> 招聘单位将会在数据库中查找关键字以便管理 // 所以你的简历应该含有一些关键字 // 例如通讯 // 网络 // 工程等 ///α<410521>

关键字一般有三种类型 // 名词 // 形容词和一些工作中用到的专业词汇  
///a<417630> 注意 // 对关键字的使用应做到尽量准确 ///a<422242> (2 // 添加你相关的优势 ///a<425037> 想一想你还有什么优势与目标职位有关 // 把它们添加到简历末尾 // 如执照 // 证书等 // 这些很可能在你的简历中起到画龙点睛的作用 ///a<437257> (3 // 重新组织 // 建立清晰合理的句子 ///a<441585> 把你的工作经历中相关内容组织在一起 // 成为段落 // 并重新排列这些段落 ///a<448366> 在句子中多加一些动词 / 以增强说服力 ///a<452529> (4 // 添加简要说明 ///a<455036> 这一点尤为重要 ///a<457340> 等一切工作都做完 // 在简历的开头部分 // 用四五句话 / 描述你的个人能力与特殊技巧 ///a<464647> 记住 {%com: end of background music} // 招聘人员只会用 10 秒左右时间来考虑一份简历是否值得他继续读下去 ///a<472515> 如果你这份简历前三分之一的部分足够吸引人 // 那么 // 你就向成功获得这个职位又迈进了一步 ///a<481378> 一份简历做好之后 // 你还可以去咨询老师以及身边有经验的同学 // 让他们给你提些参考意见 ///a<489005> 出色的简历需要有效地投递出去 ///a<491832> 下面 // 我们就给大家介绍几种有效投递简历的方式 ///a<495981> {%com: background music} 如何投递简历 ///a<499421> 一般来说 // 应聘者可以通过以下三种途径投递简历 ///a<504255> 1 // 招聘会递交简历 ///a<507173> 招聘会投递简历最大的优势在于 // 你能与招聘人员面对面沟通 ///a<513705> 不过 / 缺点也很明显 ///a<515873> 招聘会的职位大多面向有经验的人士 // 毕业生应尽量寻找对工作经验要求相对较低 // 或无明确经验要求的职位 ///a<526471> 应该注意的是 // 要充分利用招聘会现场的有利条件 // 与招聘人员积极沟通 // 尽可能多地了解该企业 // 以及该职位的信息 ///a<537561> 2 // 网络投递简历 ///a<540622> 要有针对性地挑选知名招聘网站 // 然后仔细地筛选 // 过滤信息 // 还要特别注意职位信息的有效期 ///a<549556> 找到合适的职位后 // 按照企业要求的方式进行投递 ///a<554621> 3 // 使用电子邮件投递简历 ///a<558155> 千万不要只把简历放在附件中发送出去 // 还要在信件的正文中粘贴你的简历 ///a<565882> 在邮件正文中 // 对照用人单位的要求 / 扼要地介绍你自己 // 注意使用礼貌用语 ///a<572737> 要发送标题鲜明 // 吸引眼球的应聘邮件 ///a<577611> 例如 // 具有 5 年广告公司市场部管理经验 {%com: end of background music} ///a<581955> 我们在这期节目中 // 为大家介绍了简历的三种基本格式 // 制作流程以及有效的投递方式 // 希望能对大家有所帮助 ///a<591995> 如果你想重复收看本期节目 // 请关注我们的网站 ///a<595859> 培训驿站将继续为您提供有用的求职技巧 ///a<599730> 感谢您的观看 // 我们下期再见 !

□<602347>

### Transcripción de *pinyin*<sup>13</sup>

\*CRJ: tóng xué men hǎo! jīn tiān shì wǒ men shàng kè de dì yī tiān。 wǒ xiān lái jiē shào yí xià wǒ men zhè mén kè, zěn me hhh {%act: close the door} lái shàng。 zhè yì mén kè wǒ men xué xí de shì, xīn shí yòng hàn yǔ kè běn de dì èr cè。 zhè ge xué qī wǒ men xué xí dì shí wǔ kè hhh {%act: open the book} dào dì èr shí kè。 yí gòng wǒ men xué xí liù kè shū。 ng..... měi yì kè shū jī běn shàng shì yòng liù ge xiǎo shí。 qián ..... qián sì ge xiǎo shí huò zhě wǔ ge xiǎo shí, wǒ men xué xí shēng cí、 yǔ fǎ、 kè wén hé yí bù fēn de liàn xí。 zuì hòu yí ge xiǎo shí huò zhě liǎng ge xiǎo shí, wǒ men zuò kǒu tóu de biǎo dá liàn xí。 wǒ lái jiē shào yí xià zhè yī ..... wǒ men zěn me lái shàng zhè ge kè。 ng..... shǒu xiān ne, wǒ men měi yì kè zhōng dōu shì yào xiān xué xí shēng cí。 shēng cí yǒu yì xiē cí ne, dà jiā yì jīng xué xí guò le, yì jīng rèn shi le。 zhè ge wǒ men jiù huì jiǎng de bǐ jiao shǎo yì xiē。 yǒu yì xiē zhòng diǎn de cí ne, wǒ men huì jǔ yì xiē lì zi, rán hòu ràng dà jiā duō shuō yì xiē jù zi。 ng..... rán hòu ne, shì, mmm..... jiǎng ..... jiǎng jiě kè wén。 kè wén ne, wǒ men zhǔ yào yào duō dú。 shàng kè de shí hou wǒ men méi yǒu hěn duō de shí jiān lái dú kè wén。 suǒ yǐ xū yào dà jiā xià kè yǐ hòu, fǎn fù de dú ..... dú kè wén。 zuì hǎo shì néng dá dào bèi sòng de chéng dù。 nà me shàng kè ne, wǒ men jiù dú le kè wén yǐ hòu, lǐ jiě le kè wén de yì si, huí dá yì xiē wèn tí, gēn jù kè wén huí dá yì xiē wèn tí。 rán hòu ne, wǒ men de ..... qí shí de jiǎng kè zuì zhòng yào de yí ge zhòng diǎn shì, yǔ fǎ de bù fēn。 zhè ge yǔ fǎ ne, měi yì kè dōu yǒu jǐ ge yǔ yán diǎn。 zhè jǐ ge yǔ yán diǎn ne, wǒ men huì, zhū yī lái jiǎng jiě。 rán hòu ne, zuò hěn duō de liàn xí。 [carraspea] shàng kè de shí hou, wǒ men xī wàng dà jiā jiù shì duō shuō, ng..... jiù shì tí gāo kāi kǒu lǔ。 ng..... rán hòu ne, shì kè hòu de liàn xí。 yīn wéi wǒ men zhè ge shū de kè hòu liàn xí bǐ jiao duō, suǒ yǐ wǒ men bù kě néng zài shàng kè de shí jiān, měi yì gè liàn xí dōu zuò。 wǒ huì liè chū yì xiē zhòng diǎn de liàn xí, dà jiā shàng kè zuò, huò zhě shì yǒu yì xiē liú zuò zuò yè。 měi yì kè ne, wǒ men huì yǒu yí ge huà tí。 zhè ge huà tí tāo lùn ne, shì xū yào dà jiā tí qián zhǔn bèi。 bǐ rú shuō wǒ bù zhì gěi nǐ yí ge zuò yè。 nǐ, liǎng ge tóng xué huò zhě yí ge tóng xué, zhè ge wǒ men gēn jù měi yì kè bù tóng de qíng kuàng, lái jué dìng。 liǎng ge tóng xué yì zǔ, huò zhě shì yí ge tóng xué dú

<sup>13</sup> Para facilitar el cálculo de la frecuencia de fonemas, hemos quitado todas las marcas de la transcripción.

lì wán chéng。lái jiù zhè ge huà tí xiě yí ge xiǎo de wén zhāng, huò zhě yí ge duì huà。rán hòu shàng kè de shí hòu ne, nǐ xū yào dào qián biān lái biǎo yǎn。zhè ge zhǔ yào shì wèi le liàn xí dà jiā de kǒu tóu biǎo dá néng lì。ng..... guān yú ..... zhè shì guān yú wǒ men shàng kè de bù zhòu。nà me guān yú kǎo shì, shì zhè yàng de。wǒ men de kǎo shì ne, fēn chéng sān ge bù fēn。dì yī ge bù fēn ne, shì bǐ shì。zhè ge, ng..... zhàn bǎi fēn zhī qī shí。dì èr bù fēn shì kǒu shì, zhàn bǎi fēn zhī èr shí。dì sān bù fēn jiù shì píng shí de zuò yè, zhè ge zhàn bǎi fēn zhī shí。suǒ yǐ, rú guǒ, ng..... nǐ píng shí de zuò yè bù jiāo de huà, nà me, nǐ zuì hòu de chéng jì kě néng jiù bú huì tè bié hǎo。ér wǒ jué de xué xí zhōng wén ne, yīn wéi dà jiā de ..... xiàn zài zài xī bān yá de zhè ge yǔ yán huán jìng lǐ, kě néng méi yǒu tài duō de jī huì lái shuō hàn yǔ。dàn shì, wǒ jué de wǒ men kě yǐ tōng guò wǎng luò ne, lái kàn yì xiē zhōng wén de diǎn yǐng, ng..... huò zhě shì wǒ men zhè ge kè wén ne, yě yǒu zhè ge CD。dà jiā kě yǐ yì biān ..... yì biān zǒu lù a, huò zhě yì biān zuò bié de shì qing de shí hòu, kě yǐ yì biān tīng zhè ge kè wén de CD, huò zhě shì tīng yì xiē zhōng wén gē qǔ。huò zhě yǒu shí jiān de shí hòu, kàn yì diǎn zhōng wén diǎn yǐng。zhè yàng duì yú tí gāo tīng lì, shì hěn yǒu bāng zhù de。zài yǒu yí ge ne, wǒ xī wàng dà jiā shàng kè de shí hòu, jìn liàng duō shuō hàn yǔ。bāo kuò kè jiān ..... ng..... kè jiān xiū xi de shí hòu, rú guǒ wǒ men xué guò de jù zi, kě yǐ shuō de jù zi, jiǎn dān de jù zi, jìn liàng yòng hàn yǔ lái biǎo dá。hái yǒu ne, jiù shì xū yào yù xí hé fù xí。zhè ge shì xū yào tè bié qiáng diào de。jiù shì ..... yīn wéi wǒ men de shí jiān, měi yí kè ..... měi yí ge xīng qī zhǐ yǒu sì jié ..... sì ge xiǎo shí de hàn yǔ kè。běn lái shí jiān jiù hěn duǎn。rú guǒ nǐ bú fù xí bú yù xí, hěn róng yì jiù wàng diào le。ng..... zhè shì ..... wǒ zài jiǎng yí xià zhè běn shū ..... tā de zhè běn shū de.....zhè ge.....zhǔ yào de ān pái de xiǎng fǎ, jiù shì tā de zhǔ yào de xiǎng fǎ。jiù shì zhè yí kè ..... zhè yì běn shū ne, cóng ..... zhè ge ..... cóng zhè ge ng..... zuò zhě lái biān xiě zhè běn shū de shí hòu ne, tā kǎo lǜ de yǒu zhè me jǐ ge fāng miàn。dì yī ge ne, jiù shì shuō nǐ zài yǔ fǎ jiào xué de jiē duàn, guān yú zhòng diǎn de jù xíng, hái yǒu zhòng diǎn de cí yǔ。zhè ge wǒ men jiù jué de, chéng nián rén de xué xí dì èr zhǒng ..... chéng nián rén lái xué xí dì èr zhǒng yǔ yán, ng..... bù wán ..... bù kě néng wán quán shì yì zhǒng zì rán de xí dé。zhǐ shì máng mù de mó fǎng huò zhě sǐ jì yìng bèi, kě néng dá bú dào hěn hǎo de xué xí xiào guǒ。suǒ yǐ ne, bì xū yào duì kè wén dāng zhōng chū xiàn de cí yǔ hé jù xíng jiā yǐ kòng zhì, lái tōng guò yí dìng de guī zé, ng..... zǔ cí zǔ jù de yì xiē guī zé, lái dá dào zhǎng wò yǔ yán de mù dì。tè bié shì xī bān yá yǔ hé hàn yǔ ne, yǒu hěn bù tóng de tè diǎn。yǒu ..... yǒu hěn míng xiǎn de bù tóng

de tè diǎn。suǒ yǐ ne, jiù xū yào zhè ge xué xí de shí hòu, yǒu yí dìng de xì tǒng xìng, lái yǐn dǎo xué shēng ..... yǐn dǎo xué xí zhě ne, yǒu zhèng què de zǒng jié cí yǔ zǔ jù de yí ge guī zé, lái jìn xíng jiāo jì。ng..... zhè ge shì zài ..... shì yīng gāi bǐ jiào qiáng diào de yí diǎn。jiù shì shuō, xué xí de ..... xué xí hàn yǔ de shí hòu, yào ..... xū yào àn zhào yí dìng de guī zé, hàn yǔ de guī zé ..... yǔ yán guī zé lái ..... lái xué xí。bù néng shì suí biàn de bǎ yí xiē cí yǔ zǔ zài ..... zǔ zài yí qǐ。zhè yang xué chū lái de yǔ yán ne jiù shì bù zhǔn què de huò zhě ..... huò zhě shuō bù zhā shí de。suǒ yǐ wǒ men zhè běn shū de měi yí kè de zhè ge zhòng diǎn ne, dōu shì ..... měi yí kè kè wén lǐ chū xiàn de zhòng diǎn dōu shì, tǐ xiàn zài yǔ fǎ de bù fēn。ng..... zhèi xiē yǔ fǎ, zài kè wén dāng zhōng ne, jiù huì yǒu duō cì de chóng fù, chóng xiàn。suǒ yǐ ne, jiù shì yí qián xué guò de zhòng diǎn jù xíng hé yǔ fǎ, zài zhè yí kè zhōng, kě néng yě huì duō cì de chóng fù。suǒ yǐ, zài ..... suǒ yǐ wǒ qiáng diào dà jiā, fǎn fù de duō dú kè wén, duō bèi kè wén, yě jiù shì zài zhè yí fāng miàn。rú guǒ nǐ néng gòu bǎ kè wén hěn shù liàn de shuō chū lai, nà me qí zhōng de zhè ge yǔ yán diǎn, nǐ yě jiù zhǎng wò le。ng..... hǎo, hái yǒu ne, jiù shì zhè ge liàn xí hé ..... liàn xí de bù fēn ..... ng..... liàn xí de bù fēn ne, wǒ men xū yào, ng..... yǒu zhòng diǎn de jìn xíng ..... ng..... jìn xíng zhè ge xuǎn zé。zhǔ yào shì wéi rào zhe zhè yí kè zhòng yào de cí huì hé yǔ fǎ。ng..... hái yǒu ne, jiù shì zhè ge shēng cí。shēng cí de bù fēn, měi yí ..... zhè yí ..... měi yí kè de shēng cí ne bǐ jiào duō, dà gài yǒu sì shí ge zuǒ yòu。suǒ yǐ ne, yào zhù yì, xuǎn zé zhòng diǎn de cí yǔ lái zhǎng wò。rú guǒ yí xiē bú shì hěn cháng yòng de, huò zhě shì shū miàn yǔ ..... shū miàn yǔ sè cǎi tè bié qiáng de cí yǔ, kě yǐ zàn shí bú yòng ..... bú yòng zhǎng wò ..... ng..... hái yǒu hàn zì, duì, zhè ge duì ng..... xī bān yá de tóng xué lái shuō ne yě shì yí ge nán diǎn。jiù shì hàn zì shì ..... hàn zì yīng gāi zěn me xué? zhè ge wǒ men zài zhè běn ..... zhè běn ..... zhè běn shū ne tā měi yí kè de nà ge liàn xí cè shàng ..... xué shēng de liàn xí cè shàng, tā dōu yǒu hàn zì de liàn xí。nǐ kě yǐ mó fǎng zhe zhè ge shū zhōng liè chū de, zhèi xiē hàn zì lái yí bǐ yí huà de xué ..... xué xí。rú guǒ hàn zì bù néng ng..... zhè ge kuài sù de ..... bù néng zhè ge zhǎng ..... bù néng zhǎng wò yí dìng de hàn zì de huà, nǐ zài jīn hòu xué xí hàn yǔ de guò chéng zhōng, jiù huì yù dào hěn duō de kùn nán。zài bāo kuò yuè dú shàng, hái yǒu zhè ge ng..... lǐ jiě fāng miàn, huì yù dào hěn duō de kùn nán。suǒ yǐ bù néng shuō yīn wéi hàn zì duì dà jiā lái shuō bǐ jiào nán, jiù fāng qì le。bì xū jiān chí xiě hàn zì hé rèn dú hàn zì。ng..... kǒu yǔ tīng lì de shuǐ píng ne, zhè ge, gāng cái ng..... wǒ shuō, jiù shì tīng lì ne, wǒ shuō kě yǐ dà jiā yòng kè wài de shí jiān, shì dàng de kàn yí xiē zhōng



guó de diàn yǐng, huò zhě shì tīng zhōng guó gē qǔ, huò zhě tīng wǒ men kè wén de zhè ge ng..... CD。 hái yǒu ne, jiù shì kǒu yǔ de fāng miàn ne, nǐ kě yǐ ..... yīn wéi xiàn zài wǒ zhī dào yǒu hěn duō zhōng guó liú xué shēng zài zhèr。 suǒ yǐ dà jiā rú guǒ yǒu ..... yǒu kě néng de huà, kě yǐ zhǎo yì xiē zhōng guó liú xué shēng, rèn shi yì xiē zhōng guó liú xué shēng。 píng shí ne, gēn tā men liáo tiānr。 nǐ ..... nǐ men ..... nǐ kě yǐ zhè ge shuō xī bān yá yǔ, rán hòu tā shuō hàn yǔ。 zhè zhǒng zì rán de liáo tiān de guò chéng ne, yě shì hěn hǎo de yí ge xué xí fāng shì。 hái yǒu wǒ jué de bèi kè wén yě shì yí ge hǎo bàn fǎ。 jiù shì shuō bèi kè wén ne yě shì xué xí yǔ yán de yí ge hǎo bàn fǎ。 jiù shì nǐ tōng guò bèi kè wén ne, zhè xiē yǔ yán diǎn, kě yǐ zì rán de jì zài nǐ de nǎo zi lǐ, ér qiě nǐ huì péi ..... nà ge péi yǎng chū zhèng què de yí ge yǔ gǎn。 ng..... zhè ge néng gòu bāng zhù nǐ zì rán liú lì de biǎo dá, ér bù xū yào yòng yí ge yǔ yán guī zé, zhè yang qù táo tā。 zhè ge ..... ng..... zhè ge shì wǒ men de jiào cái。 dāng rán hái yǒu yì xiē qí tā de cān kǎo jiào cái。 ng..... wǒ ..... wǒ kě yǐ gào su dà jiā, dàn shì wǒ bù zhī dào nǐ men kě bù kě yǐ zài xī bān yá zhǎo dào, kě néng bú shì hěn róng yì。 wǒ yǐ ..... wǒ yǐ jīng cóng hàn bàn dīng le yì xiē jiào cái。 dàn shì yīn wéi shì tā men yào hǎi yùn, suǒ yǐ bǐ jiào cháng de shí jiān, dà gài yí ge yuè yǐ hòu, cái kě néng yùn dào。 zhè ge xiàn zài ne, zhōng guó yě chū le yì xiē zhēn duì xī bān yá yǔ guó jiā de jiào cái, jiù shì tā de zhù shì fān yì dōu shì yòng xī bān yá yǔ de。 ng..... qí zhōng wǒ zhī dào yǒu yí ge bǐ jiào hǎo de jiào cái, wǒ jué de tā hěn kǒu yǔ huà de ne, jiù shì yí ge běi dà chū de jiào cái, jiào 《bó yǎ hàn yǔ》。 nà ge ne, tā shì yīng wén zhù shì。 tā shì cóng zuì jī chǔ de líng qǐ diǎn, yì zhí dào zuì gāo jí de ng..... zhè ge shuǐ píng dōu yǒu。 zhè ..... zhè ge jiào cái ne, tā duì yǔ fǎ de jiě shì bǐ jiao jiǎn dān míng què。 ng..... ér qiě tā de zhè ge yǔ yán ne, shì hěn dì dào de jiǎn dān de kǒu yǔ。 zhè ge wǒ jué zhe duì ..... tè bié shì duì chū xué zhě lái shuō, nà běn shū shì bú cuò de yì běn shū。 ng..... hái yǒu ne, rú guǒ děng wǒ de zhè ge jiào cái sòng lái yǐ hòu, hái yǒu bāo kuò yì xiē guāng pán, wǒ kě yǐ gěi dà jiā kàn yí xià。 rán hòu dà jiā ng..... rú guǒ yǒu xìng qu de huà, kě yǐ xiǎng bàn fǎ mǎi dào tā。 duì, yīn wéi hěn duō tóng xué yào qù zhōng guó, yǐ hòu hǎo xiàng wǒ men xì yǒu zhè zhǒng jì huà。 jiù shì shuō nǐ sì nián de xué xí dāng zhōng, zhì shǎo yào yǒu yì nián qù zhōng guó liú xué de shí jiān。 zhè zhǒng ..... zhè zhǒng qíng kuàng ne, wǒ jiù xū yào ..... wǒ jiù xī wàng dà jiā néng gòu zài zhōng guó de shí hòu, ng..... jìn liàng de bú ..... bú guò duō de gēn shuō xī yǔ de rén lái lái jiē chù, ér shì jìn liàng de gēn wài guó rén, jiù shì xī bān yá yǔ ..... ng..... bú shì ..... mǔ yǔ bú shì xī bān yá yǔ de zhè ge rén lái lái jiē chù。 yīn wéi nǐ zài zhōng guó de nà ge

huán jìng zhōng, nǐ shàng kè de bān lǐ de tóng xué kě néng hěn duō dōu .....  
dōu shì lái zì bù tóng guó jiā de, nà me duō gēn tā men jiē chù。 hái yǒu ne jiù  
shì duō gēn zhōng guó rén jiē chù。 nǐ qù shāng chǎng, nǐ qù mǎi ..... chāo shì  
mǎi dōng xī, nǐ jìn liàng de yòng nǐ xué guo de yǔ yán, xué guo de hàn yǔ lái  
biǎo dá。 zhè shì yì fāng miàn。 hái yǒu ne, jiù shì ..... jiù shì jìn liàng tiào chū  
nǐ xī bān yá yǔ de zhè ge shēng huó quān zi, lái duō jiē chù qí tā de yǔ ..... qí tā  
guó jiā de, yòng hàn yǔ lái ..... lái jìn xíng jiāo ..... jiāo jì。 ng..... zhì yú zhè ge  
shēng huó xí guān fāng miàn ne, què shí zhōng guó gēn xī bān yá yǒu hěn duō  
de bù tóng。 bǐ rú shuō ng..... dāng rán zuì ..... zuì cháng jiàn de, yòng kuài zi,  
yòng chā zi de zhè ge bù tóng。 hái yǒu ne, chī fàn de zhè ge fēng ..... xī sù。  
bǐ rú shuō ng..... zhè ge zhōng guó rén dōu shì yì qǐ lái chī fàn, wǒ men yì ge  
dà de pán zi lǐ de cài, wǒ men ng..... dà jiā yì qǐ chī。 hái yǒu ne, yǒu yì ge hěn  
míng xiǎn de bù tóng。 zài xī bān yá de shí hòu, kě néng chī fàn de shí hòu, rén  
men suǒ shǒu zhǐ、 tiǎn shǒu zhǐ zhè shì hěn zhèng cháng de。 dàn shì, zài zhōng  
guó rú guǒ nǐ zhè yang zuò ne, kě néng jiù huì ..... bié rén jiù huì jué de bù xí  
guān。 ng..... jiù shì shuō zhèi xiē, ng..... cān zhuō shàng de zhè ge, ng.....  
[chasquido] yì xiē xí guān。 zài yǒu ne, jiù shì bǐ rú shuō zài fàn guǎn lǐ, ng.....  
zài fàn guǎn lǐ yǒu shén me bù tóng? diǎn cài。 duì duì [asiente], zhè ge shì xū  
yào zhù yì de。 yì ge shì shuǐ。 wǒ men de zì lái shuǐ shì cóng shuǐ lóng tóu chū  
lái de shuǐ, shì bù néng zhí jiē hē de, xū yào shāo kāi。 huò zhě shì ne, nǐ yào  
mǎi nà zhǒng ..... cóng chāo shì lǐ mǎi chéng píng de kuàng quán shuǐ。 hái yǒu  
yì ge ne, jiù shì hěn duō xǐ shǒu jiān, kě néng méi yǒu wèi shēng zhǐ。 yě xū yào  
nǐ zì jǐ dài。 zhè ge yào zhù yì。 o duì, zhōng guó de zhè ge gōng gòng qì chē  
shén me dōu shì yǒu shòu piào yuán de。 yì bān lái shuō shì xū yào nǐ mǎi piào。  
ér qiě tā shì gēn jù nǐ zuò de zhè ge jù lí de cháng duǎn, ng..... lái shǒu bù tóng  
de qián。 jià gé bù yí yàng。 jiù shì shuō qiáng diào ..... wǒ men qiáng diào tīng  
shuō dú xiě, zhè sì gè fāng miàn。 dàn shì shí jì shàng, zuì zhòng yào de hái shì  
tīng hé shuō。 suǒ yǐ shàng kè ne, jiù, zuì qiáng diào de jiù shì kāi kǒu lǔ, ng.....  
qiáng diào jīng jiǎng duō liàn。 lǎo shī shǎo jiǎng, rán hòu xué shēng duō shuō。  
dàn shì duì yú chū jí bān lái shuō, hàn zì yě ..... yě fēi cháng zhòng yào。 hàn zì  
de rèn dú hé ..... hé shū xiě yě fēi cháng zhòng yào。 hǎo, wǒ men lái kàn .....  
wǒ men lái kàn dì shí wǔ kè ba hhh {%act: open the book}。 jīn tiān wǒ men  
xué xí dì shí wǔ kè。 nà me zhè yì kè shū zhōng zuì zhòng yào de yì ge yǔ fǎ jiù  
shì qíng tài bǔ yǔ。 zhè ge bǔ yǔ duì wài guó xué shēng lái shuō, shì bǐ jiào nán  
de yì ge yǔ fǎ。 yīn wéi hěn duō yǔ yán dāng zhōng, méi yǒu zhè ge bǔ yǔ。 bǐ  
rú shuō ng..... wǒ bā diǎn lái xué xiào le。 lái de zǎo ma? lái de hěn zǎo。 zhè ge

de, jiù shì lái de zǎo zhōng zhè ge jiù shì wǒ men suǒ shuō de, bǔ ..... qíng tài bǔ yǔ. ng..... nà me dì èr ge ..... ng..... hái yǒu yí ge fāng miàn ne, jiù shì shuō zài yòng ..... zhè shì dì yī ge bǔ yǔ de yòng fǎ. qíng tài bǔ yǔ dì yī ge yòng fǎ jiù shì méi yǒu bīn yǔ, jiù zhí jiē shuō. lái de hěn zǎo. jīn tiān nǐ jǐ diǎn lái xué xiào de? jīn tiān wǒ qī diǎn jiù lái le. tā ..... nǐ lái de hěn zǎo. nà me dì èr zhōng qíng kuàng ne, hái yǒu yí zhōng ..... yí zhōng qíng kuàng. jiù shì, shàng hǎi rén shuō pǔ tōng huà shuō de hǎo bù hǎo ? nà me wǒ men shuō, shàng hǎi rén shuō pǔ tōng huà, nà wǒ yòng qíng tài bǔ ..... yòng zhè ge bǔ yǔ de shí hòu, wǒ men bì xū yào zài shuō, shàng hǎi rén shuō pǔ tōng huà shuō de hěn hǎo, bì xū zài shuō yí biàn zhè ge shuō. yào chóng fù yí biàn zhè ge ..... zhè ge dòng cí. líng wài ne wǒ men yě kě yǐ zhè yàng shuō, ng..... pǔ tōng huà shàng hǎi rén shuō de hěn hǎo, bǎ zhè ge pǔ tōng huà zhè ge cí ne, fāng dào qián biān qù. zhè ge shí hòu, wǒ men jiù bù xū yào zài chóng fù zhè ge dòng cí. zhè shì dì yī ge zhòng diǎn yǔ fǎ. dì èr ge yǔ fǎ ne shì le. zhè ge le ne, shì biǎo ..... shì fāng zài jù zì de zuì hòu biān. biǎo shì yǐ qián zuò le yí ge shì qíng, zhè ge shì qíng yǐ jīng wán le, zhè ge shì qíng yǐ jīng wán chéng le, zuò wán le. zhù yì ..... xū yào zhù yì de jiù shì shuō, zhè ge le ne, tā yí dìng yào fāng zài jù zì de zuì hòu. yǐ qián wǒ men xué guò de zhè ge le ne, shì fāng zài dòng cí de hòu mian. bǐ rú shuō wǒ kàn le ..... zuó tiān wǒ kàn le yí ge diàn yǐng. zhè ge le ne shì fāng zài dòng cí de hòu mian. zuó tiān wǒ kàn le yí ge diàn yǐng. dàn shì wǒ men jīn tiān yào xué de zhè ge ..... zhè ge le ne, zhè ge shì fāng zài jù zì de zuì hòu. bǐ rú shuō, tā qù shàng hǎi le. tā qù shàng hǎi le. nà me zhè ge jù zì de yì sī ne, jiù shì tā yǐ jīng qù shàng hǎi le. zhè ge shì qíng yǐ jīng wán le. tā qù zhè ge dòng zuò yǐ jīng zuò wán le. dì sān ge yǔ fǎ, zhè ge bǐ jiào jiǎn dān, jiù shì, zhè ge ng..... dòng cí de chóng dié. bǐ rú shuō wǒ men shuō, xué xí ..... xiū xi, xiū xi. wǒ men kě yǐ shuō, wǒ lèi le, wǒ xiǎng xiū xi xiū xi. zhè ge dòng cí de chóng fù ne jiù biǎo dá, shí jiān hěn duǎn, huò zhě shì bú shì hěn dà de, bú shì hěn zhòng yào de yí ge shì qíng, hěn jiǎn dān de yí ge shì qíng. zhè ge shí hòu, wǒ men kě yǐ yòng chóng fù zhè ge dòng cí. zài bǐ rú shuō, zhè ge cài hěn hǎo chī, nǐ cháng chang. nà me zhè ge cháng chang ne, jiù shì biǎo shì shì yí shì de yì sī. ng..... wǒ zài shuō shuō guān yú fǎ yīn de wèn tí. hàn yǔ de fǎ yīn ne, duì ng..... xī bān yá de tóng xué lái shuō, shì hěn nán de. wǒ jué de zuì nán de jiù shì tā de shēng diào. ng..... zhè ge ā á ǎ à sì ge shēng diào dāng zhōng, yǒu qí nán de shì èr shēng hé sì shēng. yě jiù shì èr shēng shàng bú qù. zǒng shì.....bǐ rú shuō, ní hǎo. zhè ge dì yī ge nǐ biàn chéng le ní hǎo, cháng cháng shuō ní hǎo, zhè ge shēng yīn shàng ..... èr shēng shàng bú qù. hái yǒu ne, jiù

shì sì shēng xià bù lái。 sì shēng ne dú de yě bǐ jiào píng。 zhè ge shì yí ge zhǔ yào wèn tí。 jiù shì èr shēng yào wǎng shàng qù, ér sì shēng ne yào xià jiàng。 bǐ rú shuō, wǒ cóng xué yuán, qù yóu jú。 zhè ge yóu jú liǎng ge zì dōu shì, èr shēng。 ng..... tā cóng yīng guó lái zhōng guó, lái zhōng guó。 tè bié shì èr shēng hé sì shēng fāng zài yì qǐ liǎng ge cí, zài yì qǐ de shí hòu, kě néng jiù huì chū xiàn zhè yàng de wèn tí, jiù shì dú de bù zhǔn què。 bǐ rú shuō, xué xiào, yí ge èr shēng, yí ge sì shēng。 nà me zhè ge xū yào dà jiā, jiù yù dào zhè zhǒng cí de shí hòu, yào tè bié de zhù yì。 xué xiào, èr shēng yào tí shàng qu, ér zhè ge sì shēng ne, yào jiàng xià lái。 zhè shì yí ge shēng diào fāng mian de wèn tí。 zài yǒu yì xiē ..... yí ge wèn tí ne, jiù shì yǒu yì xiē fā yīn。 zài xī bān yá yǔ lǐ shì méi yǒu de。 bǐ ..... bǐ rú, jiù shì yì xiē bào pò yīn。 bǐ rú shuō b hé p, g hé k zhè yàng de。 g hé k, zhè ge k, yào yǒu qì ..... qì liú lái ..... chū lái。 zhè ge shí hòu nǐ yào yòng yì diǎnr lì qì。 zhè ge liǎng ge ..... zhè ..... zhèi xiē yīn de zhè ge qū bié。 hái yǒu jiù shì z c s hé zhè ge j q x zhè jǐ ge yīn。 j q x。 zhè jǐ ge yīn, shì bú shì zài xī bān yá yǔ yě méi yǒu ? j q x yě méi yǒu? hái yǒu z c s yě méi yǒu ? ng..... suǒ yǐ zhè jǐ ge yīn ne, xué xí qǐ lái shì bǐ jiào kùn nán de。 xū yào dà jiā duō tīng, tīng le yǐ hòu ne mó fǎng, yào jiū zhèng yǐ qián shuō xī bān yá yǔ de yí ge ..... yì xiē fā yīn xí guǎn。

## Conversión a formato XML

```
<?xml version="1.0" encoding="UTF-8" ?>
- <Transcription>
- <Header>
  <Title>mascota</Title>
  <File>chin13</File>
- <Participants>
- <Speaker>
  <ShortName>JJJ</ShortName>
  <Sex Type="woman" />
  <Age Type="A" />
  <Education Type="3" />
  <Origin>student</Origin>
  </Speaker>
- <Speaker>
  <ShortName>ZHF</ShortName>
```

```

<Sex Type="man" />
<Age Type="A" />
<Education Type="3" />
<Origin>student</Origin>
  </Speaker>
  </Participants>
<Date>01/05/2009</Date>
<Place>Alcala de Henares</Place>
<Situation>in a room, not hidden</Situation>
<Topic>Rabbits, dogs, cats, ducks</Topic>
<Source>C_ORAL_CHINA</Source>
<Class Type1="informal" Type2="public" Type3="dialogue" />
<Length>16' 20"</Length>
<Words />
<Acoustic_quality Type="A" />
<Transcriber>Dong Yang</Transcriber>
<Revisor>AZY</Revisor>
<Comments />
  </Header>
- <Text>
  <UNIT speaker="JIJ" startTime="0" endTime="2.067">哎 / 你养了一只兔子 //
    是吧 ? </UNIT>
  <UNIT speaker="ZHF" startTime="0" endTime="3.639">是啊 hhh {%act: assent}
    // 你怎么知道的 ? </UNIT>
  <UNIT speaker="JIJ" startTime="3.639" endTime="5.733">&mm 地球人都知道
    ///</UNIT>
  <UNIT speaker="ZHF" startTime="5.733" endTime="5.992">hhh {%act: laugh}
    ///</UNIT>
  <UNIT speaker="JIJ" startTime="5.992" endTime="7.961">那 [/] 那个兔子是什
    么种类的 ? </UNIT>
  <UNIT speaker="ZHF" startTime="7.961" endTime="9.101">&mm 安哥拉兔
    ///</UNIT>
  <UNIT speaker="JIJ" startTime="9.101" endTime="10.425">什么样子呢 ?
    </UNIT>

```

<UNIT speaker="ZHF" startTime="10.425" endTime="17.118">&mm 白白的 //  
 &mm 耳朵也比较长 // 但是比其他的那种兔类 // 看起来 // 就要短些啦  
 // 因为它的毛比较长 ///</UNIT>

<UNIT speaker="ZHF" startTime="17.118" endTime="21.293">&mm 看起来像  
 一个 // &mm 大毛球一样 // 是白色的 ///</UNIT>

<UNIT speaker="ZHF" startTime="21.293" endTime="23.505">然后 // 还是笨  
 笨的样子吧 ///</UNIT>

<UNIT speaker="JIJ" startTime="23.505" endTime="25.689">那肯定很可爱 //  
 手感肯定不错 ///</UNIT>

<UNIT speaker="JIJ" startTime="25.689" endTime="27.147">那养兔子麻烦  
 吗？</UNIT>

<UNIT speaker="ZHF" startTime="27.147" endTime="30.062">&mm 其实不如  
 我想象中的麻烦 ///</UNIT>

<UNIT speaker="ZHF" startTime="30.062" endTime="34.202">&mm 因为一开  
 始 / 买的时候 // 他们会说养兔子会有很多问题 ///</UNIT>

<UNIT speaker="ZHF" startTime="34.202" endTime="41.471">但是最后 / 看起  
 来的话 // 其实每天给它收拾两次左右吧 // 就能基本保证 / 这个它的房  
 间的卫生了 ///</UNIT>

<UNIT speaker="JIJ" startTime="41.471" endTime="43.055">那它吃什么呢？  
 </UNIT>

<UNIT speaker="JIJ" startTime="43.055" endTime="43.882">很麻烦吗？  
 </UNIT>

<UNIT speaker="ZHF" startTime="43.882" endTime="46.621">&mm 我的兔子  
 是什么都可以吃 ///</UNIT>

<UNIT speaker="ZHF" startTime="46.621" endTime="49.113">它 [///] 一开始  
 // 因为兔子比较小吗 ///</UNIT>

<UNIT speaker="JIJ" startTime="49.113" endTime="51.261"><hhh {%act:  
 assent}></UNIT>

<UNIT speaker="ZHF" startTime="49.113" endTime="51.261">[<] <所以说> /  
 喂养的时候 // 比较用心 ///</UNIT>

<UNIT speaker="ZHF" startTime="51.261" endTime="57.001">然后要注意给它  
 的饮水 // 不能给它喝生水 // 要让它喝矿泉水啊 / 纯净水啊 ///</UNIT>

<UNIT speaker="ZHF" startTime="57.001" endTime="61.726">然后 // 吃的方  
 面 // 只能给它喂那个 / 精装的兔粮 ///</UNIT>

<UNIT speaker="ZHF" startTime="61.726" endTime="66.792">但是后来等兔子  
 到三个月以后 // 然后 / 我就开始试着给它喂各种各样的东西 ///</UNIT>

<UNIT speaker="ZHF" startTime="66.792" endTime="72.8">比如说 / 水啊 //  
 &mm 比如说胡萝卜啊 / 苹果啊 // 然后各种蔬菜 ///</UNIT>

<UNIT speaker="ZHF" startTime="72.8" endTime="81.075">然后它特别喜欢那  
 种 / 咬起来口感有点儿硬 // 而且口感比较甜的食物 // 比如说 / 胡萝卜  
 // 这是它最喜欢的/// 然后苹果也很喜欢 ///</UNIT>

<UNIT speaker="ZHF" startTime="81.075" endTime="93.125">其他的 / 像蔬菜  
 // 豆类 // 然后 // &mm 玉米 // 马铃薯 // 土豆儿 / 都可以吃 [ ] 噢不 /  
 马铃薯之类之类都可以吃 ///</UNIT>

<UNIT speaker="JIJ" startTime="93.125" endTime="96.494">你实在太专业 /  
 太体贴了 // 你们家兔子肯定很幸福 ///</UNIT>

<UNIT speaker="JIJ" startTime="96.494" endTime="101.285">&mm 那 // 我记  
 得上次 / 我给它喂了块苹果 // 它都不吃了 ///</UNIT>

<UNIT speaker="JIJ" startTime="101.285" endTime="102.64">就是被你喂馋了  
 吧 ? </UNIT>

<UNIT speaker="ZHF" startTime="102.64"  
 endTime="104.437"><&mm></UNIT>

<UNIT speaker="JIJ" startTime="102.64" endTime="104.437">[<] <你太宠着它  
 了> hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="104.437" endTime="105.415">&mm 这倒不  
 是的 ///</UNIT>

<UNIT speaker="ZHF" startTime="105.415" endTime="109.179">因为上次你过  
 来的时候 // 是 / 下午的时候 ///</UNIT>

<UNIT speaker="ZHF" startTime="109.179" endTime="112.447">然后我们众所  
 周知 // 兔子跟我们人类的时间差 // 是很大的 ///</UNIT>

<UNIT speaker="ZHF" startTime="112.447" endTime="114.947">一般它的活动  
 时间是在晚上十点之后 ///</UNIT>

<UNIT speaker="JIJ" startTime="114.947" endTime="115.089">&mm  
 ///</UNIT>

<UNIT speaker="ZHF" startTime="115.089" endTime="118.659">有次 [ ] 上次  
 你过来的时候 // 它正在 [ ] 相当 [ ] 相当于它正在睡觉 ///</UNIT>

<UNIT speaker="ZHF" startTime="118.659" endTime="121.831">然后 // 先 [ ]  
 &mm 比较难跟它交流 ///</UNIT>

<UNIT speaker="ZHF" startTime="121.831" endTime="126.866">但是一般到晚  
 上三四点钟的时候 // 它就会特别活跃 // 这个时候是它正常的进餐时间  
 ///</UNIT>

<UNIT speaker="JJJ" startTime="126.866" endTime="128.656">兔子叫什么  
呢 ? </UNIT>

<UNIT speaker="ZHF" startTime="128.656" endTime="131.089">&mm 是西语  
名还是中文名 ? </UNIT>

<UNIT speaker="JJJ" startTime="131.089" endTime="132.289">当然是西语名了  
///</UNIT>

<UNIT speaker="ZHF" startTime="132.289" endTime="133.517">&mm 它叫  
LECHUGA ///</UNIT>

<UNIT speaker="JJJ" startTime="133.517" endTime="136.283">&mm 那它现在  
认识你吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="136.283" endTime="140.726">&mm 我觉得  
它现在只认识胡萝卜 // 并不认识它的主人 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="JJJ" startTime="140.726" endTime="146.9">我就觉得像兔子这  
种脑袋比较小的动物 // 都是脑容量不足 // 所以它肯定不认识主人  
///</UNIT>

<UNIT speaker="JJJ" startTime="146.9" endTime="148.484">这就跟狗不一样  
///</UNIT>

<UNIT speaker="JJJ" startTime="148.484" endTime="151.615">像狗 / 头那么  
大 // 所以它就好比兔子要聪明很多 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="151.615" endTime="153.526">是吗 / 你原来  
难道养过狗 ? </UNIT>

<UNIT speaker="JJJ" startTime="153.526" endTime="155.582">我没有养过狗 //  
但我特别喜欢狗 ///</UNIT>

<UNIT speaker="ZHF" startTime="155.582" endTime="157.168">是吗 ?  
</UNIT>

<UNIT speaker="JJJ" startTime="157.168" endTime="158.931">因为我妈不准我  
养狗 ///</UNIT>

<UNIT speaker="JJJ" startTime="158.931" endTime="161.51">她就是 // 所有带  
毛的动物她都害怕 ///</UNIT>

<UNIT speaker="JJJ" startTime="161.51" endTime="163.825">所以我就不能养  
狗 // 但是我特别喜欢狗 ///</UNIT>

<UNIT speaker="ZHF" startTime="163.825" endTime="165.189">hhh {%act:  
laugh} 你为什么不能养一只乌龟呢 ? </UNIT>

<UNIT speaker="JJJ" startTime="165.189" endTime="167.388">我不喜欢乌龟啊  
///</UNIT>



<UNIT speaker="JIJ" startTime="167.388" endTime="171.518">因为我 [/] 我不喜欢那种就是不认识主人的 // 然后头特别小的动物 ///</UNIT>

<UNIT speaker="JIJ" startTime="171.518" endTime="175.278">像狗就脑容量比较大 // 就比较聪明 // 所以我比较喜欢 ///</UNIT>

<UNIT speaker="JIJ" startTime="175.278" endTime="180.334">而且像那种大狗 / 让人觉得比较有安全感 // 又毛毛的 // 摸上去比较舒服 ///</UNIT>

<UNIT speaker="ZHF" startTime="180.334" endTime="181.703">&mm 你刚才提到了大狗 ///</UNIT>

<UNIT speaker="JIJ" startTime="181.703" endTime="183.583"><&mm></UNIT>

<UNIT speaker="ZHF" startTime="181.703" endTime="183.583">[<] <那> 你最喜欢哪种大型犬啊 ? </UNIT>

<UNIT speaker="JIJ" startTime="183.583" endTime="184.999">哈士奇 / 哈士奇 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="184.999" endTime="185.823">哈士奇是吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="185.823" endTime="186.57">雪橇犬 ///</UNIT>

<UNIT speaker="ZHF" startTime="186.57" endTime="188.187">就是那种灰白相间的 ///</UNIT>

<UNIT speaker="JIJ" startTime="188.187" endTime="190.226">对 hhh {%act: assent} 有蓝眼睛的那种 ///</UNIT>

<UNIT speaker="ZHF" startTime="190.226" endTime="192.25">噢 / 它是产自哪个国家的呀 ? </UNIT>

<UNIT speaker="JIJ" startTime="192.25" endTime="195.83">爱斯基摩人的吧 hhh {%act: laugh} // 我不知道 ///</UNIT>

<UNIT speaker="JIJ" startTime="195.83" endTime="199.291">因为它本 [///] 它不是又叫哈士奇又叫雪橇犬 ///</UNIT>

<UNIT speaker="JIJ" startTime="199.291" endTime="201.277">所以本来是用来拉雪橇的 ///</UNIT>

<UNIT speaker="JIJ" startTime="201.277" endTime="206.022">可能就是在很寒冷的地方 // &mm 用来拉雪橇的那种狗 ///</UNIT>

<UNIT speaker="JIJ" startTime="206.022" endTime="207.427">很可爱 / 很可爱 ///</UNIT>

<UNIT speaker="ZHF" startTime="207.427" endTime="210.361">&mm 确实狗是我们人类最好的伙伴 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="JJJ" startTime="210.361" endTime="214.353">因为第一次见哈士奇 // 是在谢霆锋一张专辑的封面上 ///</UNIT>

<UNIT speaker="JJJ" startTime="214.353" endTime="217.376">他放着那只哈士奇 hhh {%act: laugh} // 我就觉得特别可爱 ///</UNIT>

<UNIT speaker="JJJ" startTime="217.376" endTime="220.246">然后后来 [/] 后来又看了一个电影 // 叫南极大冒险 ///</UNIT>

<UNIT speaker="JJJ" startTime="220.246" endTime="226.651">就是讲 // 好多只哈士奇一起在 [/] 就是南极的那种历险故事 // 就很好玩儿 ///</UNIT>

<UNIT speaker="JJJ" startTime="226.651" endTime="231.362">然后咱们学校以前有好多只哈士奇 // 我都抱过 // 很好玩儿 ///</UNIT>

<UNIT speaker="ZHF" startTime="231.362" endTime="234.357">&mm 你说到咱们学校 // 我想起来了 // 原来咱们学校还有不少野猫 ///</UNIT>

<UNIT speaker="ZHF" startTime="234.357" endTime="235.542">你还记得它们吗 ? </UNIT>

<UNIT speaker="JJJ" startTime="235.542" endTime="238.677">当然了 // 有时候还去喂它们 ///</UNIT>

<UNIT speaker="ZHF" startTime="238.677" endTime="239.138">是吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="239.138" endTime="244.254">我记得有一只叫大黄的 // 然后胖胖的 // 已经被我们学校的女孩子宠得很挑剔了 ///</UNIT>

<UNIT speaker="JJJ" startTime="244.254" endTime="246.849">噢 / 那只大黄就很像加菲 ///</UNIT>

<UNIT speaker="JJJ" startTime="246.849" endTime="249.246">长得很像加菲猫是吧 ? </UNIT>

<UNIT speaker="ZHF" startTime="249.246" endTime="250.192"><&mm></UNIT>

<UNIT speaker="JJJ" startTime="249.246" endTime="250.192">[<] <就> 肥肥的那个 ///</UNIT>

<UNIT speaker="ZHF" startTime="250.192" endTime="250.748">对 hhh {%act: assent} ///</UNIT>

<UNIT speaker="JJJ" startTime="250.748" endTime="252.264">简直就是明星脸 ///</UNIT>

<UNIT speaker="JJJ" startTime="252.264" endTime="253.075">对啊对啊 ///</UNIT>

<UNIT speaker="JIJ" startTime="253.075" endTime="260.308">他们说 // 那只猫 // 现在 / hhh {%act: click} 由于我们学校的 / 那个同学们比较友爱 // 已经连 / 火腿肠都不吃了 ///</UNIT>

<UNIT speaker="JIJ" startTime="260.308" endTime="263.464">要给它吃鱼香肉丝之类的成菜 // 它才有胃口 ///</UNIT>

<UNIT speaker="JIJ" startTime="263.464" endTime="266.151">看来动物的命运也真是 / 各不相同 ///</UNIT>

<UNIT speaker="JIJ" startTime="266.151" endTime="269.585">有的动物这么幸福 // 有的动物那么可怜 ///</UNIT>

<UNIT speaker="JIJ" startTime="269.585" endTime="272.34">有的猫连水煮鱼都吃过 ///</UNIT>

<UNIT speaker="ZHF" startTime="272.34" endTime="273.45">hhh {%act: laugh} 就是就是 hhh {%act: assent} ///</UNIT>

<UNIT speaker="ZHF" startTime="273.45" endTime="275.165">那你原来家里养过什么动物吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="275.165" endTime="283.332">我养过小鸡 / 小鸭 / 小鱼 // 还有 [///] 反正都是 / 就没有一个能活得长的 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="283.332" endTime="285.647">鸭子看起来是很可爱的动物呢 ? </UNIT>

<UNIT speaker="JIJ" startTime="285.647" endTime="288.073">没错 // 但是是被我淹死的 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="288.073" endTime="289.494">hhh {%act: laugh} // 淹死的 ? </UNIT>

<UNIT speaker="JIJ" startTime="289.494" endTime="296.154">因为好像它是只旱鸭子 // 那我把它抱到水里面 // 让它游泳 // 然后它就 / 可能最后游不动了 // 就 [/] 就不行了 ///</UNIT>

<UNIT speaker="ZHF" startTime="296.154" endTime="297.85">是啊是啊 // 我家原来也养过鸭子 ///</UNIT>

<UNIT speaker="ZHF" startTime="297.85" endTime="300.235">然后 / 是一只很贪吃的鸭子 ///</UNIT>

<UNIT speaker="ZHF" startTime="300.235" endTime="304.857">然后 / 周末的时候 // 我会和我爸爸妈妈一起去公园里 // 给它捞鱼 ///</UNIT>

<UNIT speaker="ZHF" startTime="304.857" endTime="307.724">用网兜 / 网小鱼给它吃 // 它特别喜欢 ///</UNIT>

<UNIT speaker="JJJ" startTime="307.724" endTime="312.369">我觉得你 // 你对宠物都特别好 // 特别有爱心 ///</UNIT>

<UNIT speaker="JJJ" startTime="312.369" endTime="314.231">很专业  
///</UNIT>

<UNIT speaker="ZHF" startTime="314.231" endTime="315.37">谢谢谢谢  
///</UNIT>

<UNIT speaker="ZHF" startTime="315.37" endTime="319.201">&mm 我们家原来还 [/] hhh {%act: click} [/] &mm 我们 [/] 我们家原来养过兔子  
///</UNIT>

<UNIT speaker="ZHF" startTime="319.201" endTime="321.381">就是 [///] 但是是一只野兔 ///</UNIT>

<UNIT speaker="JJJ" startTime="321.381" endTime="323.897"><&mm></UNIT>

<UNIT speaker="ZHF" startTime="321.381" endTime="323.897">[<] <那时候>是我和我爸爸去打乒乓球 ///</UNIT>

<UNIT speaker="ZHF" startTime="323.897" endTime="325.043">在一个乒乓球  
=</UNIT>

<UNIT speaker="ZHF" startTime="325.043" endTime="327.689">我们因为当时是是在一个学校大院里打 ///</UNIT>

<UNIT speaker="ZHF" startTime="327.689" endTime="329.859">然后在乒乓球台下面发现它的 ///</UNIT>

<UNIT speaker="ZHF" startTime="329.859" endTime="333.318">然后 // 是一只比较野性难驯的兔子吧 ///</UNIT>

<UNIT speaker="ZHF" startTime="333.318" endTime="338.634">来我们家一开始 // 两个 [/] 两个多月基本上不太吃东西 // 然后奄奄一息的样子  
///</UNIT>

<UNIT speaker="ZHF" startTime="338.634" endTime="346.957">后来时间长了 // 相处得比较熟了 // 然后 // &mm 它才开始渐渐地吃东西 // 开始渐渐地阳台上活动起来 ///</UNIT>

<UNIT speaker="ZHF" startTime="346.957" endTime="348.804">最后还做了一个小草窝 ///</UNIT>

<UNIT speaker="JJJ" startTime="348.804" endTime="350.198">&mm 很好  
///</UNIT>

<UNIT speaker="JJJ" startTime="350.198" endTime="350.813">你知道吗 ?  
</UNIT>

<UNIT speaker="JIJ" startTime="350.813" endTime="354.986">那个 / 其实最烈性的动物是 [/] 是那个麻雀 ///</UNIT>

<UNIT speaker="JIJ" startTime="354.986" endTime="357.005">就是它永远不会被驯服 ///</UNIT>

<UNIT speaker="JIJ" startTime="357.005" endTime="361.221">如果你把一个麻雀弄 [/] 弄过来养的话 // 它就会绝食 // 直到 [/] 直到死 ///</UNIT>

<UNIT speaker="JIJ" startTime="361.221" endTime="368.518">因为我小时候有一次下大雨 // 住平房 // 然后我爸就在那个屋檐底下 // 就捡了一只小麻雀回来 ///</UNIT>

<UNIT speaker="JIJ" startTime="368.518" endTime="372.142">但就养到第二天 // 想等它好点把它放走 // 结果它就死了 ///</UNIT>

<UNIT speaker="JIJ" startTime="372.142" endTime="374.199">好像是自己咬断舌头自尽了 ///</UNIT>

<UNIT speaker="JIJ" startTime="374.199" endTime="376.354">因为麻雀就特别有性格 ///</UNIT>

<UNIT speaker="ZHF" startTime="376.354" endTime="380.467">hhh {%act: surprise} 噢 // 还有这 [/] hhh {%act: laugh} 这么有性格的动物 ///</UNIT>

<UNIT speaker="JIJ" startTime="380.467" endTime="386.503">就是我 [/] 我也很奇怪 // 就后来看到一个 [/] 看篇文章才知道 // 原来麻雀这么烈性 ///</UNIT>

<UNIT speaker="ZHF" startTime="386.503" endTime="389.1">&mm ///</UNIT>

<UNIT speaker="JIJ" startTime="389.1" endTime="392.744">但是我最喜欢的动物还是狗 ///</UNIT>

<UNIT speaker="ZHF" startTime="392.744" endTime="393.621">是吧 ? </UNIT>

<UNIT speaker="JIJ" startTime="393.621" endTime="394.125">&mm hhh {%act: assent} ///</UNIT>

<UNIT speaker="JIJ" startTime="394.125" endTime="395.355">你喜欢小狗吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="395.355" endTime="398.349">&mm 挺喜欢的 ///</UNIT>

<UNIT speaker="ZHF" startTime="398.349" endTime="401.822">然后 [///] 但是 // 因为狗的种类有很多吗 ///</UNIT>

<UNIT speaker="ZHF" startTime="401.822" endTime="406.846">&mm 要 [///] 我比较喜 [/] 我比较喜欢 =</UNIT>

<UNIT speaker="ZHF" startTime="406.846" endTime="409.583">我都不知道 //  
 有一种叫吉娃娃的狗 // 你知道吗 ? </UNIT>

<UNIT speaker="JJJ" startTime="409.583" endTime="410.529">知道  
 ///</UNIT>

<UNIT speaker="JJJ" startTime="410.529" endTime="411.552">我不喜欢那个狗  
 ///</UNIT>

<UNIT speaker="ZHF" startTime="411.552" endTime="415.038">对啊 // 我觉  
 得很可爱的 // 长的像 ET // 长得像外星人一样 ///</UNIT>

<UNIT speaker="JJJ" startTime="415.038" endTime="419.672">但是那种小狗 //  
 就是你让它放在家里面跑 / 一不小心 // 就会踩着它 ///</UNIT>

<UNIT speaker="ZHF" startTime="419.672" endTime="419.821">hhh {%act:  
 laugh} ///</UNIT>

<UNIT speaker="JJJ" startTime="419.821" endTime="420.798">太可怕了  
 ///</UNIT>

<UNIT speaker="ZHF" startTime="420.798" endTime="422.309">噢 / 对 // 你  
 听说过吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="422.309" endTime="424.428">一般狗越小脾  
 气越大 ///</UNIT>

<UNIT speaker="JJJ" startTime="424.428" endTime="425.432">没错  
 ///</UNIT>

<UNIT speaker="JJJ" startTime="425.432" endTime="428.937">像那种大狗 //  
 一般你去摸它 // 它都没反应 // 都很好 // 很温顺 ///</UNIT>

<UNIT speaker="JJJ" startTime="428.937" endTime="432.177">然后有时候小狗  
 // 你一上前 // 它就会乱叫 ///</UNIT>

<UNIT speaker="ZHF" startTime="432.177" endTime="433.573">对对对 hhh  
 {%act: assent} ///</UNIT>

<UNIT speaker="JJJ" startTime="433.573" endTime="437.445">我有一次在那个  
 / 银行门口 // 看一只狗被栓在那儿 ///</UNIT>

<UNIT speaker="JJJ" startTime="437.445" endTime="439.8">它主人可能进去取  
 钱了 // 它在那栓着 ///</UNIT>

<UNIT speaker="JJJ" startTime="439.8" endTime="445.473">然后它嘴上就戴着  
 那个 [/] 就是那个 / 护套还是什么 // 就是防止它咬人的那种东西  
 ///</UNIT>

<UNIT speaker="ZHF" startTime="445.473" endTime="445.756">&mm  
 ///</UNIT>

<UNIT speaker="JJJ" startTime="445.756" endTime="449.709">我当时就觉着它很可怜/// 因为那么小 // 又没什么攻击性 // 又戴上那个 ///</UNIT>

<UNIT speaker="JJJ" startTime="449.709" endTime="452.347">然后后来我就想 / 想过去摸摸它 ///</UNIT>

<UNIT speaker="JJJ" startTime="452.347" endTime="455.594">然后我 / 还 [/] 就刚走了一步 // 它就冲我特别凶地叫 ///</UNIT>

<UNIT speaker="JJJ" startTime="455.594" endTime="458.104">然后我终于知道为什么它要戴那个护套了 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="458.104" endTime="462.494">&mm 但是我  
觉得 // 在养动物的过程中 // 我们还是应该比较尊重动物的天性的  
///</UNIT>

<UNIT speaker="JJJ" startTime="462.494" endTime="462.711">&mm  
///</UNIT>

<UNIT speaker="ZHF" startTime="462.711" endTime="466.93">比如说 / 原来  
有的人养狗的时候 // 他会给狗 / 戴上嘴套或者头套 ///</UNIT>

<UNIT speaker="ZHF" startTime="466.93" endTime="470.99">就是因为狗会叫  
// 然后叫的话显得比较不乖吗 // 有失主人的身份 ///</UNIT>

<UNIT speaker="ZHF" startTime="470.99" endTime="474.172">但是长此以往  
的话 // 这个狗就会特别胆小 // 非常怯懦 ///</UNIT>

<UNIT speaker="ZHF" startTime="474.172" endTime="478.74">然后跟同伴 //  
就是它在街上遇到别的狗的时候 // hhh {%act: click} 之间交流也会有问题  
///</UNIT>

<UNIT speaker="ZHF" startTime="478.74" endTime="483.75">然后 // 这时候  
我觉得狗 // 应该 // 相当不开心的 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="JJJ" startTime="483.75" endTime="484.587">&mm 没错  
///</UNIT>

<UNIT speaker="JJJ" startTime="484.587" endTime="489.179">还有 // 我觉得  
狗本来就是应该是家养动物 // 因为是从狼驯化来的 ///</UNIT>

<UNIT speaker="JJJ" startTime="489.179" endTime="493.337">所以就是 // 有  
时候看到街上那些流浪狗 // 就觉得它们很可怜 ///</UNIT>

<UNIT speaker="JJJ" startTime="493.337" endTime="496.956">因为它们已经失  
去了动物的天性 // 没办法自己在野外生存 ///</UNIT>

<UNIT speaker="JJJ" startTime="496.956" endTime="498.515">就那些人太不负  
责任了 ///</UNIT>

<UNIT speaker="JJJ" startTime="498.515" endTime="501.489">你既然把它驯养  
了 // 就不应该抛弃它 ///</UNIT>

<UNIT speaker="ZHF" startTime="501.489" endTime="502.075">&mm  
 ///</UNIT>

<UNIT speaker="JIJ" startTime="502.075" endTime="505.315">因为那只是你生命中的几年 // 但是 / 是小狗的一辈子 ///</UNIT>

<UNIT speaker="ZHF" startTime="505.315" endTime="506.464">对对对对对  
 hhh {%act: assent} ///</UNIT>

<UNIT speaker="ZHF" startTime="506.464" endTime="509.638">&mm 你这么喜欢狗 // 你喜欢猫吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="509.638" endTime="510.213">不喜欢  
 ///</UNIT>

<UNIT speaker="ZHF" startTime="510.213" endTime="510.97">为什么呀 ?  
 </UNIT>

<UNIT speaker="JIJ" startTime="510.97" endTime="514.585">因为猫头很小  
 hhh {%act: laugh} // 所以脑容量不足 // 跟兔子是一个原理 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="514.585" endTime="517.511">看起来你喜欢各种各样头大的动物 hhh {%act: laugh} // 是吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="517.511" endTime="519.144">不是头大 / 就是比较聪明的 ///</UNIT>

<UNIT speaker="ZHF" startTime="519.144" endTime="519.89">认主人 ?  
 </UNIT>

<UNIT speaker="JIJ" startTime="519.89" endTime="520.237">对 ///</UNIT>

<UNIT speaker="JIJ" startTime="520.237" endTime="527.443">我觉得猫有时候 // 就是 [///] 反正不是很 [/] 不是很认主人 // 而且比较冷 // 不像狗那么忠心 ///</UNIT>

<UNIT speaker="ZHF" startTime="527.443" endTime="533.667">&mm 有一种说法 // 说狗比较像男性 // 比较忠诚 // 然后猫比较像女性 // 比较有性格 ///</UNIT>

<UNIT speaker="ZHF" startTime="533.667" endTime="536.69">然后 / 有时候你冷落它 // 它 [/] 它也会不理你 ///</UNIT>

<UNIT speaker="ZHF" startTime="536.69" endTime="541.62">然后 // &mm 比较难以猜测它的心思吧 ///</UNIT>

<UNIT speaker="JIJ" startTime="541.62" endTime="548.091">&mm 我觉得 / 就猫 / 就是那种 / 没有表情的动物 // 但是狗就是有表情的 ///</UNIT>

<UNIT speaker="ZHF" startTime="548.091" endTime="550.093">狗是有表情的  
 hhh {%act: surprise} ? </UNIT>



<UNIT speaker="JIJ" startTime="550.093" endTime="554.945">我觉得是 // 因为它 [ ] 它要是难过 // 它会 // 就那种呜呜地叫 ///</UNIT>

<UNIT speaker="JIJ" startTime="554.945" endTime="558.5">然后 / 开心的话 // 就会 [///] 我觉得狗是会笑的 ///</UNIT>

<UNIT speaker="ZHF" startTime="558.5" endTime="560.797"><&mm></UNIT>

<UNIT speaker="JIJ" startTime="558.5" endTime="560.797">[<] <我不> 知道你  
能不能看出来 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="JIJ" startTime="560.797" endTime="562.676">反正这是 / 我的  
感觉 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="562.676" endTime="563.47">是吗 ?  
</UNIT>

<UNIT speaker="ZHF" startTime="563.47" endTime="567.751">&mm 那根据  
你的了解 // 狗一般是怎么跟人类交流的呢 ? </UNIT>

<UNIT speaker="JIJ" startTime="567.751" endTime="572.543">就是 // 用 //  
摇尾巴呀 // 来表示友好 ///</UNIT>

<UNIT speaker="JIJ" startTime="572.543" endTime="576.993">然后 / 用一些  
就各种各样的叫声 // 来表达它自己的想法 ///</UNIT>

<UNIT speaker="JIJ" startTime="576.993" endTime="579.787">然后 [///] 或者  
可以拖着你玩啊 // 什么的 ///</UNIT>

<UNIT speaker="JIJ" startTime="579.787" endTime="583.834">因为我没养过狗  
// 所以我不了解 // 我都是到处找别人的狗玩儿 ///</UNIT>

<UNIT speaker="ZHF" startTime="583.834" endTime="588.229">hhh {%act:  
laugh} 其实 / 很多动物都有自己的语言 ///</UNIT>

<UNIT speaker="ZHF" startTime="588.229" endTime="592.861">比如说 // 蜜  
蜂发现食物的时候 // 它会圆圈 [///] 原地跳八字舞 ///</UNIT>

<UNIT speaker="ZHF" startTime="592.861" endTime="597.699">然后当兔子感  
觉危险的时候 // 它会不断蹬后蹄发出响声 // 来警告自己的同类  
///</UNIT>

<UNIT speaker="ZHF" startTime="597.699" endTime="602.508">然后虽然有些  
宠物 / 现在已经家养了 // 但是它们原始的天性还是得以保留的  
///</UNIT>

<UNIT speaker="JIJ" startTime="602.508" endTime="604.354">&mm 很有意思  
///</UNIT>

<UNIT speaker="ZHF" startTime="604.354" endTime="610.719">&mm 你知道  
藏獒吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="610.719" endTime="612.99">&mm 知道 // 很有个性的狗 ///

<UNIT speaker="ZHF" startTime="612.99" endTime="617.5">&mm 在我国的西北部 // 藏獒有 [/] 有一种说法 ///

<UNIT speaker="ZHF" startTime="617.5" endTime="620.493">叫做 // &mm 一獒可敌三狼 ///

<UNIT speaker="ZHF" startTime="620.493" endTime="623.814">就是说獒是一种很忠心 // 同时很勇猛的动物 ///

<UNIT speaker="ZHF" startTime="623.814" endTime="627.679">它们只认识从 [/] 从小生下来就照顾它们的那些主人 ///

<UNIT speaker="ZHF" startTime="627.679" endTime="632.699">然后 / 很难跟它们建立关系 // 但是 / 你一旦跟它认识之后 // 它就对你 [/] 会对你特别忠诚 ///

<UNIT speaker="ZHF" startTime="632.699" endTime="638.522">然后 / 是享誉我 [/] &mm 海外 [/] 享 [/] 享誉国内外的一种很有名的犬类 ///

<UNIT speaker="JIJ" startTime="638.522" endTime="639.446">而且很臭 ///

<UNIT speaker="JIJ" startTime="639.446" endTime="643.534">因为 / hhh { %act: laugh } 我有一个好朋友他舅舅就专门养大型犬的 ///

<UNIT speaker="ZHF" startTime="643.534" endTime="643.936">&mm ///

<UNIT speaker="JIJ" startTime="643.936" endTime="650.854">然后他就 [///] 他家后院儿一个 [///] 就好多铁笼子 // 然后有几十条那种大狗 // 就藏獒还有其他别的狗 // 我忘了 ///

<UNIT speaker="JIJ" startTime="650.854" endTime="655.644">反正就 / 真的超级超级臭 // 臭得难以想象的那种 ///

<UNIT speaker="JIJ" startTime="655.644" endTime="657.686">然后而且很凶狠凶 ///

<UNIT speaker="JIJ" startTime="657.686" endTime="660.889">但是它们对它的主人就特别忠心 // 特别好 ///

<UNIT speaker="ZHF" startTime="660.889" endTime="661.564">&mm ///

<UNIT speaker="JIJ" startTime="661.564" endTime="667.855">而且我觉得现在藏獒因为 / 越来越多人把它作为家养的宠物 // 它有 [/] 已经有点儿失去它原来的天性了 ///

<UNIT speaker="JJJ" startTime="667.855" endTime="670.482">现在一只家养的藏獒肯定敌不过三头狼 ///</UNIT>

<UNIT speaker="ZHF" startTime="670.482" endTime="672.248">hhh {%act: laugh} 是啊是啊 hhh {%act: assent} ///</UNIT>

<UNIT speaker="JJJ" startTime="672.248" endTime="674.745">其实我觉得 =</UNIT>

<UNIT speaker="JJJ" startTime="674.745" endTime="676.598">你 [/] 你喜欢狼吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="676.598" endTime="681.74">&mm 我不喜欢 // 但是我比较欣赏这种动物 // 因为原来看过一本书 / 叫狼图腾 ///</UNIT>

<UNIT speaker="JJJ" startTime="681.74" endTime="681.872">&mm ///</UNIT>

<UNIT speaker="ZHF" startTime="681.872" endTime="682.948">我已经记不清作者了 ///</UNIT>

<UNIT speaker="ZHF" startTime="682.948" endTime="685.693">但是当时是正 [/] 我正在上高中的时候 ///</UNIT>

<UNIT speaker="ZHF" startTime="685.693" endTime="688.06">但这本书 / 给我留下了很深刻的印象 ///</UNIT>

<UNIT speaker="ZHF" startTime="688.06" endTime="690.103">它向我们诠释了一些概念 ///</UNIT>

<UNIT speaker="ZHF" startTime="690.103" endTime="694.545">然后 // hhh {%act: click} &mm 当时是 [/] &mm 当时对我影响很大 ///</UNIT>

<UNIT speaker="ZHF" startTime="694.545" endTime="696.558">主要是因为当时正在高考吗 ///</UNIT>

<UNIT speaker="ZHF" startTime="696.558" endTime="699.338">然后 // 人与人之间的竞争比较激烈 ///</UNIT>

<UNIT speaker="ZHF" startTime="699.338" endTime="702.017">然后那本书就向我们诠释一种狼的精神 ///</UNIT>

<UNIT speaker="ZHF" startTime="702.017" endTime="704.897">比如说在受伤的情况下不断拼搏 ///</UNIT>

<UNIT speaker="ZHF" startTime="704.897" endTime="711.143">在失去同伴的情况下 // &mm 依然能够 [/] &mm 能够自己独立地生存 ///</UNIT>

<UNIT speaker="ZHF" startTime="711.143" endTime="716.295">然后 [///] 但是同时它又非常依赖于群体 // 然后对群体有很高的信任 ///</UNIT>

<UNIT speaker="ZHF" startTime="716.295" endTime="720.126">有些狼甚至会牺牲自己的性命 // 来保全这个团 [/] 团队的利益 ///</UNIT>

<UNIT speaker="JJJ" startTime="720.126" endTime="728.146">&mm 我记得有个说法是 // 如果一只狼 / 被兽夹 / 夹住了脚的话 // 它为了 [/] 为了生存下去 // 它会自己咬断自己的骨头 ///</UNIT>

<UNIT speaker="ZHF" startTime="728.146" endTime="729.542">对对对 hhh { %act: assent } ///</UNIT>

<UNIT speaker="JJJ" startTime="729.542" endTime="735.147">&mm 我 [/] 我其实挺喜欢狼的 // 但是是那种叶公好龙式的喜欢 ///</UNIT>

<UNIT speaker="ZHF" startTime="735.147" endTime="735.37">&mm  
///</UNIT>

<UNIT speaker="JJJ" startTime="735.37" endTime="736.634">就是离的远远的  
///</UNIT>

<UNIT speaker="JJJ" startTime="736.634" endTime="737.359">我还觉得狼  
///</UNIT>

<UNIT speaker="JJJ" startTime="737.359" endTime="738.554"><狼是挺不错的  
></UNIT>

<UNIT speaker="ZHF" startTime="737.359" endTime="738.554">[<] <hhh { %act: laugh }> ///</UNIT>

<UNIT speaker="JJJ" startTime="738.554" endTime="742.085">因为那个 [/] 因为它的外形就 / 有点像哈士奇 ///</UNIT>

<UNIT speaker="ZHF" startTime="742.085" endTime="742.748">&mm  
///</UNIT>

<UNIT speaker="JJJ" startTime="742.748" endTime="744.617">就是那 [/] 那种样子我挺喜欢的 ///</UNIT>

<UNIT speaker="JJJ" startTime="744.617" endTime="750.207">然后以前还梦想过要养只狼或 <养只豹子什么的></UNIT>

<UNIT speaker="ZHF" startTime="744.617" endTime="750.207">[<] <hhh { %act: laugh }> ///</UNIT>

<UNIT speaker="JJJ" startTime="750.207" endTime="751.474">还可以保护我 hhh { %act: laugh } ///</UNIT>

<UNIT speaker="ZHF" startTime="751.474" endTime="754.661">好像施瓦辛格养了一只老虎还是狮子 // 你知道吗 ? </UNIT>

<UNIT speaker="JJJ" startTime="754.661" endTime="756.083">不知道 hhh { %com: move the head } ///</UNIT>

<UNIT speaker="ZHF" startTime="756.083" endTime="759.169">&mm 我也是听说吧 ///</UNIT>

<UNIT speaker="ZHF" startTime="759.169" endTime="764.49">反正我觉得跟  
如此大 [J] 一个大型动物在一起 // 感觉一定会比较有意思 ///</UNIT>  
<UNIT speaker="JJJ" startTime="764.49" endTime="766.171">可是这个是合法  
的吗 ? </UNIT>  
<UNIT speaker="JJJ" startTime="766.171" endTime="767.353">在美国 ?  
</UNIT>  
<UNIT speaker="ZHF" startTime="767.353" endTime="768.868">不知道  
///</UNIT>  
<UNIT speaker="ZHF" startTime="768.868" endTime="773.388">我觉着 / 估计  
是类似 [J] 是类似一种领养的性质吧 // 就好像咱们国家大熊猫  
///</UNIT>  
<UNIT speaker="ZHF" startTime="773.388" endTime="780.697">由于经济问题  
// 咱们 / 当地的 // &mm 有关机构不能够负担它的费用的话 // 他就会  
让一些人来领养大熊猫 ///</UNIT>  
<UNIT speaker="ZHF" startTime="780.697" endTime="784.188">然后通过这种  
方式 // 来让大熊猫获得更好的生存条件 ///</UNIT>  
<UNIT speaker="JJJ" startTime="784.188" endTime="788.194">&mm 我突然想  
起来 // 就是前一阵子去希腊旅游 ///</UNIT>  
<UNIT speaker="ZHF" startTime="788.194" endTime="788.431">&mm  
///</UNIT>  
<UNIT speaker="JJJ" startTime="788.431" endTime="792.767">就发现希腊 //  
就是尤其是雅典 // 到处都是流浪狗 ///</UNIT>  
<UNIT speaker="JJJ" startTime="792.767" endTime="794.823">但是那些狗都带  
着狗牌儿 ///</UNIT>  
<UNIT speaker="JJJ" startTime="794.823" endTime="796.12">然后是一样的狗  
牌儿 ///</UNIT>  
<UNIT speaker="JJJ" startTime="796.12" endTime="801.446">所以我就很奇怪  
/// 就还 [///] 一开始还以为是那个政府雇狗来看城的 / 还是什么意思  
///</UNIT>  
<UNIT speaker="JJJ" startTime="801.446" endTime="804.805">然后后来那个旅  
店的人 / 就说那个就是流浪狗 ///</UNIT>  
<UNIT speaker="JJJ" startTime="804.805" endTime="811.581">然后我就不明白  
// 因为 [J] 就是因为那些狗都是 [J] 都是那种大型犬 // 然后都带的是一  
样的狗牌儿 ///</UNIT>  
<UNIT speaker="JJJ" startTime="811.581" endTime="813.925">我就觉得那些主  
人实在是太狠心了 ///</UNIT>

<UNIT speaker="ZHF" startTime="813.925" endTime="817.552">对啊 // 但是  
这些狗牌是谁给它们装上的呢 ? </UNIT>

<UNIT speaker="JIJ" startTime="817.552" endTime="819.437">所以就 // 是个  
谜啊 hhh {%act: laugh} ! </UNIT>

<UNIT speaker="ZHF" startTime="819.437" endTime="824.432">hhh {%act:  
laugh} 都是一样的狗牌儿 // 但是狗牌儿上有不同的名字还  
是 .....</UNIT>

<UNIT speaker="JIJ" startTime="824.432" endTime="826.346">没细看 // 不敢  
接近 hhh {%act: laugh} ///</UNIT>

<UNIT speaker="ZHF" startTime="826.346"  
endTime="827.771">&neneenenene ///</UNIT>

<UNIT speaker="JIJ" startTime="827.771" endTime="829.631">但到处都是那样  
的狗 ///</UNIT>

<UNIT speaker="JIJ" startTime="829.631" endTime="832.97">就那些历史遗迹  
里面也都是 ///</UNIT>

<UNIT speaker="ZHF" startTime="832.97" endTime="833.798">是吗 ?  
</UNIT>

<UNIT speaker="JIJ" startTime="833.798" endTime="842.205">&mm 然后就像  
给 [///] 就是 [///] 就像守城的狗一样 ///</UNIT>

<UNIT speaker="JIJ" startTime="842.205" endTime="846.312">还有咱们在意大利  
那个庞贝古城里面 // 不也有狗吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="846.312" endTime="853.528">&mm 我觉着  
那些好像是那个 / 当地的政府专门 / 雇佣过来 // 然后帮 [/] 帮助保护  
文物古迹的 ///</UNIT>

<UNIT speaker="JIJ" startTime="853.528" endTime="857.205">我有时候在想 //  
会不会是以前那些狗的灵魂 ? </UNIT>

<UNIT speaker="ZHF" startTime="857.205" endTime="858.476">hhh {%act:  
laugh} ///</UNIT>

<UNIT speaker="JIJ" startTime="858.476" endTime="864.852">就是 // 那些人  
不在了 // 但那些狗还守护在那个城市 // 所以在那徘徊来徘徊去的  
///</UNIT>

<UNIT speaker="ZHF" startTime="864.852" endTime="867.31">你说的这个让  
我想起来了赤兔马的故事 ///</UNIT>

<UNIT speaker="JIJ" startTime="867.31" endTime="868.786">&mm ///</UNIT>

<UNIT speaker="ZHF" startTime="868.786" endTime="874.159">当时不是那个  
叶公 { %alt: el dato correcto es 关公 } 走麦城之后 // 然后赤兔马 // 三日不  
吃不喝 ///</UNIT>

<UNIT speaker="ZHF" startTime="874.159" endTime="877.334">然后 // 为离  
开自己的主人吗 ///</UNIT>

<UNIT speaker="ZHF" startTime="877.334" endTime="881.457">然后 // 最终  
/ hhh { %act: click } 也算是成全了它主人的一段英明吧 ///</UNIT>

<UNIT speaker="JIJ" startTime="881.457" endTime="885.185">没错 // 马就是  
除了狗以外 / 我另外喜欢的一个动物 ///</UNIT>

<UNIT speaker="JIJ" startTime="885.185" endTime="888.559">就是很有灵性 //  
很可靠 // 而且很实用 ///</UNIT>

<UNIT speaker="ZHF" startTime="888.559" endTime="889.575">你骑过马  
吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="889.575" endTime="890.264">骑过  
///</UNIT>

<UNIT speaker="ZHF" startTime="890.264" endTime="891.484">在内蒙古 ?  
</UNIT>

<UNIT speaker="JIJ" startTime="891.484" endTime="892.125">没错  
///</UNIT>

<UNIT speaker="ZHF" startTime="892.125" endTime="893.515">感觉怎么  
样 ? </UNIT>

<UNIT speaker="JIJ" startTime="893.515" endTime="894.573">很爽 hhh { %act:  
laugh } ///</UNIT>

<UNIT speaker="ZHF" startTime="894.573" endTime="896.855">那你是怎么跟  
马交流的呢 ? </UNIT>

<UNIT speaker="JIJ" startTime="896.855" endTime="901.876">那个时候 // 很  
害怕 // 就坐着一动都不敢动 // 没怎么跟它交流 ///</UNIT>

<UNIT speaker="JIJ" startTime="901.876" endTime="904.193">下次有机会再好  
好交流 hhh { %act: laugh } ///</UNIT>

<UNIT speaker="ZHF" startTime="904.193" endTime="905.985">hhh { %act:  
laugh } 你骑的那种马是有马鞍的吗 ? </UNIT>

<UNIT speaker="JIJ" startTime="905.985" endTime="911.045">应该是吧 // 那  
种给游客骑的马都是备马鞍的 // 然后还有人牵着 ///</UNIT>

<UNIT speaker="ZHF" startTime="911.045" endTime="915.101">我觉着纵马内  
蒙古一定是一件特别有意思的事情 ///</UNIT>

<UNIT speaker="JJJ" startTime="915.101" endTime="923.513">&mm 我觉得如果 // 有 [J] 有一个自己的农场 // 然后养好多狗 // 再养好多 [J] 好多匹马 // 然后人生就完美了 ///

<UNIT speaker="ZHF" startTime="923.513" endTime="925.828">hhh {%act: laugh} 很不错的理想呢 ///

<UNIT speaker="JJJ" startTime="925.828" endTime="933.521">那个 # [///] 那你养过小狗吗 ? </UNIT>

<UNIT speaker="ZHF" startTime="933.521" endTime="935.383">&mm 没有 // 我一直挺遗憾的 ///

<UNIT speaker="ZHF" startTime="935.383" endTime="940.378">因为 / 有很多朋友都在养狗/// 然后他们讲狗的各种各样有趣的故事 ///

<UNIT speaker="ZHF" startTime="940.378" endTime="945.963">讲狗的忠诚啊 // 讲 / 小狗有很可爱的一面啊 // 之类之类的 ///

<UNIT speaker="ZHF" startTime="945.963" endTime="948.666">&mm 但是一直没有机会养 ///

<UNIT speaker="ZHF" startTime="948.666" endTime="950.897">因为在家里毕竟条件比较有限吗 ///

<UNIT speaker="ZHF" startTime="950.897" endTime="957.546">&mm 一方面是经济上的 // 买一只狗在我们国内 [///] 买一只比较好的狗在国内还是很贵的 ///

<UNIT speaker="ZHF" startTime="957.546" endTime="962.123">另外我们需要不断地为它准备饮食 // 准备各种住宅 ///

<UNIT speaker="ZHF" startTime="962.123" endTime="965.113">然后帮它挑选 // hhh {%act: click} 同伴或者配偶 ///

<UNIT speaker="ZHF" startTime="965.113" endTime="970.136">然后 [///] 另外一方面就是时间问题 // 因为咱们 // 国内大多数家长还都是上班族吗 ///

<UNIT speaker="ZHF" startTime="970.136" endTime="973.378">然后每天都有工作要忙 // 晚上回来一般都比较疲劳了 ///

<UNIT speaker="ZHF" startTime="973.378" endTime="975.951">很难保证小狗一日三餐的供应 ///

</Text>

</Transcription>



## Apéndice B

### *Listados completos de datos estadísticos*

#### *1. Datos estadísticos del subcorpus Formal*

##### a. Sexo y edad

Archivo	Sexo	Edad
chfo01	M	B
chfo02	H	B
chfo03	M	D
chfo04	M	D
chfo05	M	D
chfo06	M	D
chfo07	M	D
chfo08	M	D
chfo09	M	D
chfo10	M	D
chfo11	M	D
chfo12	M	D
chfo13	M	D
chfo14	M	D
chfo15	M	D
chfo16	M	D
chfo17	M	D
chfo18	M	D
chfo19	M	D
chfo20	M	D
chfo21	M	D

b. Duración Media de los Enunciados

<b>Archivo</b>	<b>Nº caracteres</b>	<b>Enunciados</b>	<b>Caracteres/enunciado</b>
chfo01	4349	205	21.21
chfo02	3663	144	25.44
chfo03	2096	120	17.47
chfo04	1358	76	17.87
chfo05	530	35	15.14
chfo06	1392	60	23.20
chfo07	745	47	15.85
chfo08	1011	59	17.14
chfo09	1336	79	16.91
chfo10	993	56	17.73
chfo11	1060	61	17.38
chfo12	955	52	18.37
chfo13	283	13	21.77
chfo14	238	8	29.75
chfo15	1247	46	27.11
chfo16	241	10	24.10
chfo17	313	11	28.45
chfo18	738	28	26.36
chfo19	396	25	15.84
chfo20	1472	83	17.73
chfo21	414	20	20.70
media			20.06

## 2. Datos estadísticos del subcorpus Informal

### a. Sexo y edad

Archivo	Sexo	Edad
chin01	M/M	B/A
chin02	H/M	A/A
chin03	H/M	A/A
chin04	H/M	A/A
chin05	H/M	B/A
chin06	H/H	B/B
chin07	H/M	B/B
chin08	M/M	B/A
chin09	H/M	A/A
chin10	H/M	B/B
chin11	M/M	B/C
chin12	H/M	A/A
chin13	H/M	A/A
chin14	M/M	A/A
chin15	M/M	A/A
chin16	H/M	A/B
chin17	H/M	B/B
chin18	H/M	B/A
chin19	M/M	B/A
chin20	M/H	A/B
chin21	H/M	C/B
chin22	H/M	A/A
chin23	M/M	B/A
chin24	H/M	A/A
chin25	H/H	B/A

b. Duración Media de los Enunciados

Archivo	Nº caracteres	Enunciados	Caracteres/enunciado
chin01	9291	587	15.83
chin02	1978	144	13.74
chin03	1254	67	18.72
chin04	2896	120	24.13
chin05	232	11	21.09
chin06	1779	83	21.43
chin07	1462	107	13.66
chin08	2924	149	19.62
chin09	4632	396	11.70
chin10	258	21	12.29
chin11	710	78	9.10
chin12	2018	102	19.78
chin13	4594	315	14.58
chin14	1260	65	19.38
chin15	1281	74	17.31
chin16	373	49	7.61
chin17	2301	117	19.67
chin18	444	29	15.31
chin19	3690	260	14.19
chin20	1026	100	10.26
chin21	342	43	7.95
chin22	4448	258	17.24
chin23	5610	332	16.90
chin24	2638	202	13.06
chin25	2056	116	17.72
media			15.55

c. Duración Media de los Turnos

Archivo	Turnos	Caracteres/turnos
chin01	356	26.10
chin02	66	29.97
chin03	28	44.79
chin04	40	72.40
chin05	5	46.40
chin06	26	68.42
chin07	52	28.12
chin08	54	54.15
chin09	316	14.66
chin10	15	17.20
chin11	59	12.03
chin12	31	65.10
chin13	172	26.71
chin14	30	42.00
chin15	24	53.38
chin16	45	8.29
chin17	40	57.53
chin18	16	27.75
chin19	111	33.24
chin20	80	12.83
chin21	33	10.36
chin22	153	29.07
chin23	203	27.64
chin24	144	18.32
chin25	27	76.15
media		27.98

### 3. Datos estadísticos del subcorpus Medios de Comunicación

#### a. Sexo y edad

Archivo	Sexo	Edad
chme01	M	A
chme02	M	A
chme03	M	A
chme04	M/M	A/A
chme05	M	A
chme06	M	A
chme07	H/M	A/A
chme08	M	A
chme09	M/M	A/A
chme10	H/M	A/A
chme11	M	A
chme12	M	A
chme13	M/H	A/D
chme14	H/M	B/A
chme15	M/M	A
chme16	M	A
chme17	M	A
chme18	M/M	D/A
chme19	M/M	A/A
chme20	M	A
chme21	M/M/H	A/A/A
chme22	M	A
chme23	M/M	A/A
chme24	H/M	A/A
chme25	H/M	B/A

b. Duración media de los Enunciados

Archivo	Nº caracteres	Enunciados	Caracteres/enunciado
chme01	906	49	18.49
chme02	2015	97	20.77
chme03	2860	150	19.07
chme04	1947	98	19.87
chme05	172	6	28.67
chme06	660	40	16.50
chme07	2735	128	21.37
chme08	2409	105	22.94
chme09	3648	166	21.98
chme10	272	11	24.73
chme11	390	12	32.50
chme12	304	14	21.71
chme13	2834	100	28.34
chme14	2562	108	23.72
chme15	3920	154	25.45
chme16	7335	294	24.95
chme17	1368	76	18.00
chme18	3278	164	19.99
chme19	270	13	20.77
chme20	1924	90	21.38
chme21	2851	119	23.96
chme22	949	36	26.36
chme23	2570	95	27.05
chme24	4725	197	23.98
chme25	3472	176	19.73
media			22.57

#### 4. Cuadro resumen

	Nº caracteres	Enunciados
<b>Formal</b>	24830	1238
<b>Informal</b>	59497	3825
<b>Medios de comunicación</b>	56376	2498
<b>Total</b>	140703	7561



## **Apéndice C**

*Ejemplo de documento escaneado de la firma del  
permiso de grabación*

I, .....刘畅....., at the request of the person in charge of the project carried out by the Computational Linguistics Laboratory (LLI) at the Universidad Autónoma de Madrid, whose purpose is to develop a spontaneous speech corpus made up of a collection of recordings and transcriptions, within the framework of the Doctoral Dissertation by 董柳....., give my permission and authorization for the following:

1. Recording of my voice.
2. Transcription of the recording.
3. Treatment of the sound and the transcription.
4. Dissemination and commercialization of the sound and the transcription
5. I keep the right to listen to the recordings and to deny my authorization for whatever reason that I consider pertinent.

The persons in charge guarantee the anonymity of the participants and of any person mentioned in the recording.

Date: 09.1.9

Signature: 刘畅